RAILTON'S REDUCTIVE MORAL REALISM

A thesis submitted to
Kent State University in partial
fulfillment of the requirements for the
degree of Master of Arts

By
Garrett Rauckhorst

May 2013
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... v

CHAPTER

0. INTRODUCTION ................................................................................................................. 1
  0.1 Metaethics ....................................................................................................................... 1
  0.2 The Outline .................................................................................................................... 2
  0.3 Purpose ......................................................................................................................... 3

1. RAILTON’S REDUCTIVE MORAL REALISM ................................................................. 5
  1.1 The Fact/Value Distinction .......................................................................................... 5
    1.1.1 The Argument from Rational Determinability ..................................................... 6
    1.1.2 The Argument from Internalism .......................................................................... 7
    1.1.3 The Argument from Ontological Queerness ....................................................... 9
    1.2 Railton’s Positive Account of Non-moral Value ..................................................... 11
    1.3 Explanatory Uses of Non-moral Value ................................................................... 14
    1.4 Intuitive Appeal ....................................................................................................... 15
    1.5 Normative Realism .................................................................................................. 16
    1.6 Moral Value ............................................................................................................ 17
    1.7 Explanatory Uses of Moral Value ......................................................................... 19
    1.8 Moral Realism and the Argument from Relativity ................................................. 23
    1.9 Moral Value, Rationality, and Normativity ............................................................ 24
    1.10 Conclusion ........................................................................................................... 25

2. DEREK PARFIT .................................................................................................................... 27
  2.1 Parfit’s Metaphysical and Epistemological Justification ............................................. 28
    2.1.1 Metaphysical Justification ................................................................................. 28
    2.1.2 Epistemological Justification .......................................................................... 31
    2.2 The Smoker and Agony .......................................................................................... 40
    2.3 The Argument from the Source of Reasons ......................................................... 45
    2.4 Hard Naturalism ..................................................................................................... 48
    2.5 Naturalism and the Normativity Objection ............................................................ 55
    2.6 Conclusion ........................................................................................................... 60

3. CONNIE ROSATI ............................................................................................................... 61
  3.1 Ideal Advisor and Normativity .................................................................................. 62
  3.2 Ideal Advisor and Full Information .......................................................................... 64
  3.3 Rosati’s Objection to Ideal Advisors and Normativity .......................................... 66
  3.4 Rosati’s Conceptual Impossibility Objection to Ideal Advisor and Full Information .. 71
  3.5 Does Railton Meet the Internalist Requirement ....................................................... 74
3.6 Railton and the Justificatory Requirement.................................................................77
3.6.1 Ideal Advisor’s Personality.....................................................................................78
3.6.2 Railton and the Fully Informed Person.................................................................84
3.7 Conclusion..................................................................................................................90

4. WHAT HAVE I DONE? .................................................................................................91
   4.1 Parfit.......................................................................................................................92
   4.2 Rosati ......................................................................................................................94

BIBLIOGRAPHY .............................................................................................................97
ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Michael Byron. He was hip to my many attempts at straw men, non-sequiturs, and equivocations. He was incredibly patient in tolerating my questionable mastery of English grammar. His metaethical insights were crucial in my attempt to write a cogent thesis. Any mistakes that I managed to slip by him are the fault of solely myself. Thank you.

I would like to thank my readers: Dr. Frank Ryan, Dr. Kim Garchar, and Dr. Tammy Clewell. Thanks for reading my thesis thoroughly in the short period between its completion and my defense. Thanks for all of the amazing and insightful questions that were raised at the defense.

I would like to thank Elaine Blum for helping me format the final version of this paper. Without her help this paper would never have been formatted correctly.

Lastly, I would like to thank all of my friends and family for tolerating my neglect during my work on this thesis. Thank you especially my fiancée, Sarah Rhoades.
0 INTRODUCTION

0.1 Metaethics

The field of metaethics explores peculiar questions about moral philosophy. When I make a moral claim such as, "Murder is wrong," what exactly am I claiming? This claim is a first-order ethical claim, which attempts to provide us with action guidance by telling us something about murder. Metaethics instead asks questions about first-order ethical claims. What is the “wrongness” about which I speak? Am I claiming that murder is against the law, or that I do not like it, or that it is disdained by my culture, or am advising you against it? Conversely, I might be claiming that there are certain moral properties that exist in the world, and the murder has the property of being wrong. If this were the case, then what kind of strange thing would such a property be?

There are many different views to these and many other questions in metaethics. Most of them go by fancy sounding names. Cognitivists believe that moral judgments such as, “Murder is wrong,” express facts about the world and are thus either true or false. Cognitivists disagree about whether these statements are ever true. Non-Cognitivists believe that moral judgments are an expression of a preference or an emotion or some other non-truth-apt state. Along similar lines Moral Realists believe that moral properties exist in the mind-independent world and that humans are somehow capable of interacting with them. Anti-realists disagree, believing instead that there are no mind-independent moral properties.

There is some disagreement about which features constitute the metaethical divide between naturalism and non-naturalism. Naturalists, as I understand them, believe that
moral properties are exclusively natural properties. *Non-Naturalists*, as I understand them, believe that moral properties are not exclusively natural properties. Naturalists disagree about whether moral properties are fully reducible to natural properties or are utterly unique natural properties. *Reductionists* believe that they are fully reducible to other more mundane, but possibly complex properties of our natural world. *Non-reductionists* believe that moral properties are *sui generis*.

Thus metaethics is the study of several different and difficult components or moral philosophy. Metaethics examines, *inter alia*, the semantics, metaphysics, and epistemology of regular first-order ethics. Semantics is the study of meaning. Metaethics examines what we mean or attempt to mean when making ethical claims. Metaphysics is the study of what there is and what these things are like. Metaethics examines whether moral properties exist, and if so what they consist in. Epistemology is the study of knowledge. Metaethics examines how we are able to know about moral properties, if they exist.

### 0.2 The Outline

In this work I will be defending a reductive naturalistic version of moral realism, a theory proposed by a philosopher from The University of Michigan, Peter Railton. This view holds that moral properties exist. They are reducible to other natural properties. When we make ethical claims, we are making claims that are true or false, and sometimes true. Some things really are right or wrong.

By claiming that there are moral properties, all moral realists incur both epistemological and metaphysical burdens. A moral realist must explain how we know about moral properties. A moral realist must explain what moral properties are like.
Chapter 1 will provide an explication of Railton’s theory, which will include both what makes it naturalistic and what makes this naturalism reductive.

I will be defending this theory against objections raised by Derek Parfit of the University of Oxford, and Connie Rosati of the University of Arizona. Parfit objects that Railton’s theory faces problems concerning its subjectivity and its normativity. He believes that a reductive naturalism such as Railton’s must be problematically connected to human subjects, jeopardizing the mind-independence of moral properties. In addition, naturalism does not allow for the type of normativity that Parfit desires, which is one that provides everyone with irreducibly normative reasons. Chapter 2 will consider Parfit’s views. It will begin with an explanation of Parfit’s views, and then continue with an examination of his objections. Rosati also objects that Railton’s theory fails in the normativity department. However, for her this means that it has an insufficient connection to the human motivational system and insufficiently justifies its metaphysical claims. In Chapter 3 I will examine Rosati’s objections as well as possible responses. In both Chapters 2 and 3 I will argue that Railton’s theory has the resources to answer, undermine, or deflect all of these objections and still remain plausible. In Chapter 4 I will provide a brief conclusion of what I think this paper was able to accomplish.

0.3 Purpose

I decided to write this paper on moral realism because I believe that we have good reasons to believe that moral properties exist. The fact that they do exist matters greatly. These properties help us to decide what we ought to do with our lives. I chose to write on Railton’s moral realism because it seems to me to be an incredibly plausible account of the good for a person, the best one that I have encountered to date. I find that a reductive naturalistic account of value is the best way to settle the epistemological and metaphysical
debts of realism. If I can defend this account against some of its prominent objections, I hope to have contributed in some small way towards its success.
CHAPTER 1: RAILTON’S REDUCTIVE MORAL REALISM

In his articles “Moral Realism” and “Facts and Values,” Peter Railton has proposed a version of naturalistic moral realism which attempts to solve the metaphysical and epistemological burdens which have traditionally plagued moral realism. This realism requires postulating the existence of metaphysical entities which are no more outlandish than beings for whom things matter and the fact that these beings interact with their environment in predictable and reliable ways which can positively or negatively affect these things which matter (or potentially matter) to them. Railton’s theory provides reforming definitions of moral and non-moral value that aim to show that truth-apt moral judgments and moral normativity can be preserved within a purely empirical theory.¹ I will begin with Railton’s rebuttals to the prima facie arguments against value realism. I will then lay out his substantive theory of non-moral value and moral value.

1.1 The Fact/Value Distinction

The first problem facing any potential naturalistic moral realist is the putative fact/value distinction. Some claim that facts and values are two completely different kinds of entities, neither capable of being described completely in the language of the other. Railton notes that there are three common philosophical arguments for the fact/value distinction: the argument from rational determinability, the argument from internalism, and the argument from ontological queerness.² Each of these arguments presupposes that

practical rationality is instrumental in nature. Railton accepts this presupposition, but nonetheless he finds none of the arguments to be successful.

1.1.1 The Argument from Rational Determinability

The argument from rational determinability holds that due to the instrumental nature of practical rationality, reason can dictate only the means to a certain given end for an agent and never the end as such. Thus, although any two rational agents can be made to agree on the facts simply through maximal exposure to experience and logic, values are different. Two rational agents could undergo any amount of experience and be fully logical and yet disagree on matters of value. Thus, the fact/value distinction.

Railton rejects the premise that rational agents can necessarily be made to agree on all of the facts.

From the standpoint of instrumental reason, belief-formation is but one activity among others: to the extent that we have reasons for engaging in it, or for doing it one way rather than another, these are at bottom a matter of its contribution to our ends. What it would be rational for an individual to believe on the basis of a given experience will vary not only with respect to his other beliefs, but also with respect to what he desires.

Thus, although Railton concedes that people may not be forced to agree on matters of value through exposure to experience and logic, he argues that this lack of rational determinability does not ground a distinction between facts and values. People need not agree on all matters of fact through mere experience and logic either. According to the instrumental nature of practical rationality, a given act is rational for an agent, if it promotes

---

the ends of that agent. Thus, what it is rational for one to believe will be a function of both one’s other beliefs, and one’s currently held values! If one does not share the scientific values of evidence, simplicity, explanatory power, and repeatability, one may not be able to be convinced of the facts provided by empirical science.

Thus, the argument from rational determinability fails to support the fact/value distinction. Though it is true that values may not be forced on us merely by experience and logic, neither are beliefs. Given the instrumental nature of practical rationality, the rationality of all beliefs held by any given individual depend upon both that individual’s goals and that individual’s other beliefs.

1.1.2 The Argument from Internalism

The argument from internalism also begins with the premise that practical rationality is strictly instrumental. To this it adds a very strong version of internalism: in order for something to be valuable, it must provide everyone with a reason to pursue it. The instrumental nature of rationality entails that a thing is a reason for an agent only if it will allow him to further some end which he already has.

Now it has been held by contemporary Humeans and Kantians alike that in order for a normative judgment to apply to an individual, he must have some reason -- not necessarily overriding -- to comply with it. … If this were what value judgments involved, then, given the instrumental conception of rationality (and here contemporary Humeans and Kantians

---

6 Railton later calls this the thesis that in order for something to be valuable it must provide everyone with a reason to value it, value absolutism. (Railton, "Facts and Values," p. 47).
part company), it would follow at once that value-judgments could not be factual.\(^7\)

People’s ends are contingent, and these ends may differ from person to person. Thus, if claiming that something is valuable means that it provides everyone with a reason to pursue it, then values must be different from facts. People can accept facts as true without these facts necessarily providing everyone with a reason to do something. Since rational agents may have different contingent ends, and what is rational depends on the ends that different agents actually have, we cannot know a priori\(^8\) that there is anything which necessarily provides a reason for action to all rational agents.

However, Railton notes that this conception of internalism is too strong.\(^9\) Railton does accept a version of internalism for intrinsic value. He argues that since different things are valuable for different rational agents, then in order for something to be valuable for a particular rational agent, it must provide that agent with a reason to pursue it. Just because there are no absolutely valuable things should not preclude the reality of relationally valuable things. To illuminate his thesis of relational value, Railton draws an analogy between value and nutrition.\(^10\) Whether something is a nutrient depends on the organism which is consuming it. Adult gazelles are nourished by a diet of plants. Adult lions are nourished by a diet of meat. Therefore, nutrition like value is relational. Calcium is not absolutely nutritious regardless of the organism consuming it. Nonetheless, we can

\(^7\) Railton, “Facts and Values,” pp. 45-46.

\(^8\) It seems that it could happen to be a contingent fact that there is something that provides a reason for pursuit to all rational beings. Suppose that everyone actual values being in perfect health, and there were a drug that provides everyone who takes it with a lifetime of perfect health. Then, given everyone’s contingent values, everyone would still be provided with a reason to take this drug even under this strong version of internalism.


truly say that for some organisms calcium really is nutritious, and for any given rational agent, some things really are valuable.

Thus, with the loss of value absolutism, the argument from internalism fails to support the fact/value distinction. It is true that in order for something to be intrinsically valuable for a person, it must be capable of providing that person (at least under some set of circumstances) a reason to pursue it. And it is true that what provides a person with a practical reason is a function of the contingent ends of that person. However, only if a thing of must be provide everyone with a reason, in order to be valuable for anyone do we find a global fact/value distinction. Relational value is completely compatible with both instrumental rationality and a modest form of internalism. A modest form of internalism states that in order for something to be valuable for a person it need only be capable of providing that person with a reason to pursue it. Given instrumental rationality, the mere fact something is valuable for you provides you with a reason to pursue it.

1.1.3 The Argument from Ontological Queerness

The argument from ontological queerness is the ontological variation of the argument from internalism, which is motivational. The argument from ontological queerness states that if absolutely valuable things exist, then they are metaphysically queer, i.e., problematically strange, too strange to postulate. If value were something which necessarily provided a reason for action for all rational agents regardless of their contingent desires, then this value would be a very strange thing indeed. Where would it come from, and where would it get its authority? It would have to be foisted upon the agents by either God or rationality.¹¹ For the naturalist, explanatory appeals to God are not available. Our assumptions about the instrumental nature of rationality preclude the

possibility of something necessarily providing a reason for pursuit to all rational agents (since reasons are provided by the contingent ends of rational agents). The ontological status of facts on the other hand is not problematic. We all understand what it means to say that, “It is raining outside.” The ontological status of values is not as straightforward. If we said, “It is good that it is raining outside,” in what does this goodness consist? Thus, the difference between the ontological status of facts and values grounds the fact/value distinction.

However, as in the argument from internalism, the problem here is value absolutism: if something is valuable it must provide everyone with a reason to pursue it. Instead of absolute value, the type of value which Railton proposes involves only relational value, where things are good for particular agents. Such value is much more plausible than value that must provide reasons for all agents. The ontological status of value is no more mysterious than the relationship between the psychology and biology of human beings and their environment. Value requires subjectivity, beings for whom things matter. These beings also have objective properties, in light of which certain things are actually valuable to them. An example of such an objective property would be whether, for them, a certain substance is a nutrient or a poison. This relationship is a function of their biological makeup and the chemical properties of the substance.

Without the premise of value absolutism there is no need to postulate a queer ontological status for values. Relational value, in which X is valuable for A, is no more ontologically mysterious than other facts of the world. That I am a human being, and I am hydrated by drinking water, is a fact. This fact entails that in some circumstances, drinking

---

water is good for me. Thus, the argument from ontological queerness fails to ground the fact/value distinction.

1.2. Railton’s Positive Account of Non-moral Value

Railton begins his theory of value realism with a subject, someone for whom things are capable of mattering, a human being like you or me, called Lonnie. Some things matter to Lonnie, and Railton calls those his subjective interests. Lonnie cares about many things, such as his health and happiness, his day-to-day pleasures and satisfactions, the fulfillment of his desires, etc. But Lonnie’s mere interest in or desire for a thing or outcome does not necessitate that this thing is actually a constituent of Lonnie’s good. For any particular thing in which Lonnie is interested, it makes sense for him to ask whether this interest or its object is actually good for him. As a regular human being, Lonnie finds himself in a tragically meager epistemic position. He has no way of knowing what lays in store for him in the future. Perhaps eating at McDonalds five times a week will cause him to suffer a fatal heart attack at age 50, taking many years off of his life expectancy. Perhaps he will die from something unrelated and experience no ill effects from his dietary habits. Lonnie basically has to wing it. The reason why as human beings we have to wonder whether our particular interests, desires, and goals are actually good for us is that we have no way to be certain of the consequences of pursuing these subjective interests.

Lonnie’s subjective interests have a reduction basis, i.e., we can explain his interest in something by reference to his constitution and disposition and the environment in which he does and has found himself. I enjoy listening to many types of music because my mind and body are physiologically disposed to so enjoy such sounds. A lion enjoys killing and eating a gazelle because it is likewise so constituted to enjoy that experience. However, my being now disposed to enjoy a particular thing does not entail that thing is
actually good for me. Some humans are inordinately disposed to enjoy smoking crack cocaine. However, it may be in the best interest of such smokers to attempt to rid themselves of this disposition.

In order to refer to Lonnie’s actual good, Railton describes a device that he calls Lonnie’s objectified subjective interest. Lonnie’s objectified subjective interest is constituted by what an epistemically ideal version of Lonnie, Lonnie+, would want Lonnie to want, were Lonnie+ to assume Lonnie’s unique position in the universe. Lonnie+ is Lonnie with unlimited cognitive powers, full factual and descriptive information about himself, his constitution, and his environment, the ability to fully and vividly contemplate all of his possible options and their outcomes, and free from all errors in instrumental rationality. Stipulating that Lonnie+ is to enter into the place of actual Lonnie, becoming non-idealized Lonnie, ensures that we are talking about the objectified subjective interest of Lonnie, rather than Lonnie+.

To illustrate the difference between subjective interest and objectified subjective interest, Railton tells the story of dry Lonnie. Lonnie is traveling in a foreign country and begins to feel debilitated and homesick. Unbeknownst to Lonnie, his discomfort is caused by dehydration. Craving the familiar, Lonnie seeks out a large glass of warm milk. The milk aggravates his condition and Lonnie comes to feel even worse. Were Lonnie+ to consider Lonnie’s situation, he would want Lonnie not to want milk, but instead to want hydrating liquids. The fact that Lonnie+ would want Lonnie to want hydrating liquids makes it the case that hydrating liquids are in Lonnie’s objectified subjective interest.

---

14 For instance suppose that Lonnie finds himself lost in a maze. Lonnie+ would have no interest in a map for himself, given his perfect knowledge of the maze, but may want himself to want a map were he in the position of actual Lonnie, lost in a maze.
Next Railton connects objectified subjective interest to actual objective interest. He does this by noting that the reason *why* Lonnie+ would want Lonnie to want hydrating liquids is that given Lonnie’s situation they are in his objective interest. Objective interest for a particular being is what is *actually* good for him. Every person has objective interests. The objective interest of a particular person, A, gets its objectivity from the actual facts of the world which compose the reduction basis for the objectified subjective interest of A. The reduction basis for A’s objectified subjective interest includes all of the facts about A, such as A’s constitution and environment as well as how these facts would be used by A, were A idealized so as to have perfectly vivid comprehension of these facts and the possible outcomes for A, as well as perfect instrumental rationality. Thus, it is *because* Lonnie finds himself in a position in which drinking hydrating liquids is in his best interest that explains why Lonnie+ would recommend that Lonnie want them.

In establishing the objective interest of A, or A’s actual good, Railton has described what he calls non-moral value realism.\(^\text{16}\) This non-moral value bridges the fact/value distinction as it is composed entirely of descriptive facts about the universe, able to be couched in non-evaluative terms. The value is non-moral because it is only what is good for a *particular* person, without reference to the good of others. As we shall see, non-moral value is a crucial component in his case for moral realism. First, however, I must mention some explanatory uses of this theory of non-moral value.

\(^{16}\) Railton, “Moral Realism,” p. 176. After noting that A+ wants A to want certain things because they are in A’s objective interest, Railton ceases to refer to A+’s desires for A as “objectified subjective interests” and refers to them simply as A’s objective interests. Of course A+ would want A to want some things more than others, and thus we can view A as having a spectrum of objective interests ranging from *the* good for A down to merely a good for A.
1.3 Explanatory Uses of Non-moral Value

Suppose someone else in Lonnie’s dehydration predicament, Tad, drank water rather than milk. Given this concept of objective interests we can now explain the differing fates of Lonnie who drinks milk and Tad who drinks water. Lonnie suffers discomfort and displeasure, while Tad perks up and experiences well-being. The explanation of Tad and Lonnie’s differing conditions is the fact that for Tad and Lonnie, milk was not in their objective interest, while water was in their objective interest. Suppose that Tad knew that he was dehydrated and that water is hydrating, and so accordingly selected water. Then we could describe the difference between their subsequent conditions by saying that Tad knew what was good for him, while Lonnie did not know what was good for him.

The postulation of objective interest serves the further purpose of explaining what Railton calls the wants/interests mechanism. If we actually have objective interests, then this fact helps to explain why our subjective interests evolve. They often do so to more closely match what is in our objective interest. Railton illustrates this phenomenon by supposing that dry Lonnie went into a shop to purchase milk and found that the shop was out of dairy products. Instead Lonnie spots a 7-up and gets a warm fuzzy feeling of nostalgia and decides to purchase the soda instead. He consumes it and finds that it was pretty enjoyable, so he purchases a six-pack and drinks several more sodas. In the morning his dehydration is gone and so is his malaise. Due to this fortuitous cure of his ailment, Lonnie’s future interests may be slightly altered. Suppose in the future he is again traveling abroad and becomes dehydrated. Feeling the familiar malaise he remembers how good the 7-up made him feel and again successfully treats his condition. Even

---

oblivious to his dehydration, his wants may naturally evolve to more closely match his objective interests. The evolution of person’s subjective interests to more closely mirror his objective interests is the wants/interest mechanism.

Of course such a learning mechanism is not perfect.\(^{18}\) Many other mechanisms are involved which may prevent wants from evolving to match interests. The society or culture in which one finds oneself may be geared towards training people to desire things that are detrimental to their interests. One may simply fail to undergo experiences with the positive feedback necessary to prod one into altering one’s desires. The detrimental effects of one’s desires may be minute in the short-run, but devastating in the long-run. For these and many other reasons, some wants often fail to evolve towards interests.

### 1.4 Intuitive Appeal

Railton notes that this account of non-moral good also has intuitive appeal.\(^{19}\) Our current selves may question whether our wants match what is actually good for us because our current wants are based in part on our limited epistemic condition. Not so with our idealized selves. On Railton’s theory our good is what we would want ourselves to want if we knew what we were doing. However, we must keep in mind that this is an account of the non-moral good for a person. This account does not rule out the possibility that some people’s good may be constituted by some things which we may find appalling. However, this diversity of values just goes to show that human beings come in a wide variety of initial constitutions and dispositions. The fact that a thing is non-morally valuable


\(^{19}\) Railton, "Moral Realism," p. 177.
for a person does not mean that this thing cannot be criticized on moral or aesthetic
grounds.20

What is actually good for a person will most often correspond to what allows one to
experience mental and physical well-being.21 Mental and physical well-being are what we
generally conceive of to be intrinsically valuable for almost all human beings in all
circumstances, even after deepest consideration and reflection. Thus, this is what our
idealized selves would most likely want us to want from counterfactual conditions.
Although this is a general rule of thumb it need not always be the case.22 For instance
some people’s interest may also require promotion of the well-being of others, or even
self-sacrifice to a cause.

1.5. Normative Realism

In addition, Railton’s theory also provides normative realism. Given the
instrumental nature of practical rationality, practical rationality is the efficiency of achieving
an agent’s subjective interests given his beliefs. If I already want to get to the train station
as quickly as possible and know that Route A is faster than Route B, then Route A is the
one that I rationally ought to take. However, we can now talk of practical rationality in an
extended sense as efficiency of achievement of an agent’s objective interests. Suppose an
English POW, John, is attempting to escape from the Nazis in World War II Germany.
John is attempting to get to Switzerland as quickly as possible and knows that Route A is
shorter than Route B. However, suppose further that unbeknownst to him, should he take
the quicker Route A he will be recaptured by the Nazis and shot for escaping. Should John

take the more arduous Route B he will successfully reach Switzerland. We can say that in the extended sense of rationality, John ought rationally to take Route B. Were he to come to understand that Route B was the only safe route to Switzerland, then he would recognize and respond to the normative force of the fact that he ought, in the sense of extended practical rationality, to take that route.

Although the statement, “John ought to take Route B to Switzerland,” contains the normative term, ought, this statement is reducible to purely descriptive naturalistic facts. The statement could be re-phrased as, “If John wants to achieve his objective interest, then he must take Route B to Switzerland.” This hypothetical imperative maintains the normative force of the ought statement and is made true by the facts of the situation (Route A will lead to John’s death, while Route B will lead to his safety and freedom).

1.6 Moral Value

While non-moral value concerns the interests of individuals, moral value takes into account of the interests of multiple people. Railton takes the distinctive feature of morality to be that it is impartial to the interests of particular individuals and treats the interests of all individuals equally. Having defined non-moral value, i.e., the good for each particular individual, moral value is now defined as the maximization of the non-moral value of all individuals whose value is at stake. This metaethical theory presupposes a first-order normative ethical theory, a type of consequentialism in which the value to be maximized is not simply individual pleasure/or happiness, but instead individual objective interest.

---

24 Railton, "Moral Realism," p. 190n31. Being a value pluralist Railton assumes that in some cases an individual’s objective interest may differ from his pleasure/happiness, such as in cases
Thus morality, like epistemology, becomes a type of rationality. Morality is a type of practical rationality, namely practical rationality from the point of view of all persons rather than from the point of view of your particular self. Railton calls this the social point of view. In order to determine what is rational from the social point of view, the interests of all parties affected by an action must be taken into equal consideration. Therefore, in order to impartially determine what is good from the social point of view, interpersonal comparisons of value between individuals must be possible. Railton takes this to be a plausible assumption.

I am assuming that when a choice is faced between satisfying interest X of A vs. satisfying interest Y of B, answers to the question “All else equal, would it matter more to me if I were A to have X satisfied than if I were B to have Y satisfied?” will be relatively determinate and stable across individuals under conditions of full and vivid information. Thus, what one ought morally to do is that act which would maximize the non-moral value of all persons whose value are affected by that act. The moral rightness and wrongness of any particular action will come in degrees, depending on the discrepancy between that act and the actual maximization of the non-moral value of the group of persons.

where someone may choose to make a sacrifice in order to complete some achievement or contribute to some cause.

25 Railton, "Moral Realism," p. 190n.29. He acknowledges that 'social point of view' may not be the best name for a view that takes into account all persons, since societies traditionally separate themselves from other persons in other societies.


27 Railton, "Moral Realism," p. 191. Many philosophers prefer to say that acts are either right or wrong, and that only the goodness and badness of acts come in degrees. Railton is not one of those philosophers.
1.7 Explanatory Uses of Moral Value

Railton’s theory of moral realism has some explanatory uses analogous to the explanatory uses of his theory of non-moral value.\textsuperscript{28} Those people whose subjective interests fail to track their objective interests are unlikely to experience happiness and general well-being, and less likely in proportion to the incongruence between their subjective and objective interests. In a similar spirit, in a society whose individuals fail to approximate rationality from the social point of view, one is less likely to find that society experiencing peace and prosperity. This likelihood decreases in proportion to the incongruence between the actions of its individuals and social rationality. This incongruence will probably take the form of some type of discrimination against a group, segment, or class of people in the society.

For example, consider the protests led by Dr. Martin Luther King Jr. in the southern United States during the Civil Rights movement. Railton can explain this unrest in terms of the failure of the United States to approximate social rationality. In this case discounting the interests of a significant minority led to a backlash against an unjust social structure, and the subsequent social reform was an instance of movement towards social rationality. The increase in social rationality in turn led to decrease in social unrest. In a similar fashion, moral value allows us to predict that in the future, instances in which societies diverge significantly from social rationality will tend to lead to the social unrest of their peoples.

Due to the complexities of large groups of people, these ill effects may be less obvious in the social model than in the case of a single individual. The group of people whose interests have been neglected may not notice that their interests have been so

\textsuperscript{28} Railton, "Moral Realism," p. 191.
neglected, or may not care, or may be unable to do anything about it. In other cases, a departure from what is rational from the social point of view may manifest problems such as ill-will, protest, revolt, and increasing chaos and disorder. Such unrest may occur even if the group whose interests are discounted is not explicitly aware that any injustice is occurring. The mere fact that the injustice is occurring may lead to social friction. In order for the hypothesis that departures from social rationality can have these consequences to be testable, predictions must be made about what social outcomes will occur in particular circumstances due to the relations of particular situations and the social rationality of those situations.

In the case of the non-moral good of a person, Railton postulated a wants/interests mechanism, in which the subjective interests of an individual evolve through increased experience to more closely track the objective interests of that individual. An analogous mechanism can be postulated to be at work in the case of societies and social rationality. Suppose that a particular society, S, systematically neglected the interests of a group of people, G. G may recognize this disparity in justice and stage a protest, which inconveniences everyone in the society. As a consequence S may consider G’s position and re-evaluate S’s treatment of G through the merits of G’s argument. In this fashion S’s infrastructure may evolve towards social rationality.

Again due to the complexities of large groups of people, the feedback mechanism in a society might not function as smoothly as the wants/interests mechanism of an individual. When things go wrong for individuals, they are likely to notice that something is wrong, even if they do not know exactly what is wrong. Due to the disparities in power

---

between groups of people in societies there is no guarantee that a given society will be consistently moving towards social rationality. As Railton says,

Of course, other mechanisms have been at work influencing the evolution of social practices and norms at the same time, some with the reverse effect. Whether mechanisms working on behalf of the inclusion of excluded interests will predominate depends on a complex array of social and historical factors.\(^{31}\)

The feedback mechanism for social rationality may not be the strongest mechanism at work in a situation. There may be multiple other mechanisms pushing the society in different directions. Some mechanisms which may work against social rationality may be religious or oppressive ideologies, vast differences of initial wealth, and the inherent limitations of subpopulations.\(^{32}\) Predictions based on the wants/interests mechanism social rationality will need to take into account the many other mechanisms at play in a given situation and movement towards social rationality will only occur in fairly specific conditions.

Railton admits that the truth of his postulated social rationality feedback mechanism can be confirmed only through extensive empirical investigation. However, he does mention some general historical trends that may be considered prima facie evidence in favor of the theory. The first is the expansion of the moral sphere.\(^{33}\) Initially in human history, people considered members of other tribes less morally relevant than their fellow tribesmen. The moral sphere has since increased to include more and more diverse


\(^{32}\) As examples of subpopulations with special difficulties Railton lists "children and future generations" (Railton, "Moral Realism," p.194n35).

peoples. Now in the present day, all plausible moral theories in Western civilization consider every human being to be of equal moral worth. The second is the humanization of morality. Morality has from time to time been given a variety of bizarre explanations such as supernatural origin, rules to be blindly followed, and the demands of pure reason. Now morality is grounded in the well-being of creatures for whom things are capable of mattering. The third is that there will be greater trends towards social rationality in some situations than in others. As Railton explains,

We should expect to see greater approximation in those areas of normative regulation where the mechanisms postulated here work best, for example, in areas where almost everyone has importantly similar or mutually satisfiable interests, where almost everyone has some substantial potential to infringe upon the interests of others, where the advantages of certain forms of constraint or cooperation are highly salient even in the dynamics of small groups, and where individuals can significantly influence the likelihood of norm-following behavior on the part of others by themselves following norms. In fact this is exactly what we see, i.e., it is in areas where we have mutually satisfiable interests that we find the greatest approximation of social rationality. Prohibitions against murder, theft, and wanton violence are ubiquitous across cultures. In situations where there is a great disparity in power, or mutually exclusive goals, or a great disparity in the interests of different groups, social rationality is not approximated. Given enough

---

information the social rationality feedback mechanism could both predict and explain these outcomes.\textsuperscript{36}

1.8 Moral Realism and the Argument from Relativity

Railton considers another common objection to moral realism, the argument from relativity.\textsuperscript{37} This argument points out that there is a radical diversity of moral opinions across societies, and there has been no uniform universal progress towards a moral consensus. These facts, it is claimed, make it unlikely that our moral theories are tracking an underlying moral reality. The diversity of moral opinions is said to be better accounted for by the hypothesis that morality is merely conventional and that people call things right because they happen to do them, rather than doing them because they are right.\textsuperscript{38}

However, Railton argues that this argument proves too much. It applies \textit{mutatis mutandis} against scientific realism. The fact that many cultures do not accept the findings of science does not entail that our scientific theories are merely conventional. Since the proponents of this argument are unwilling to accept it as an argument against scientific realism, they should likewise be unwilling to accept it as an argument against moral realism.

Railton’ naturalistic account of moral value provides us with a theory of moral realism which avoids the traditional pitfalls of non-naturalistic theories of moral realism. The account does not require the postulation of metaphysically mysterious entities, which are allegedly apprehended through quasi-sensory perceptions. It does not require the belief that the universe be concerned with human affairs. It does not postulate a theory of substantive practical rationality, some of which claim that pure reason is capable of

\textsuperscript{36} Railton, "Moral Realism," p. 199.
providing rational agents as such with reasons to be moral. Moral realism depends only on the existence of human beings, beings for whom things matter, and the fact that things can go objectively better and objectively worse for these beings. Given this we can say confidently and truly assert that some things are really wrong, and ought not be done, and that some things are really right, and ought to be done.

1.9 Moral Value, Rationality, and Normativity

Do the moral oughts provided by Railton’s theory of moral realism provide the necessary normative force to motivate actual agents to behave morally? What is rational from the point of view of a particular agent, and what is rational from the social point of view may fail to coincide. For instance a person may be attempting to decide whether to spend his money on a dishwasher (practically rational in the extended individual sense), and sending it to Africa to provide clean water and vaccinations for children he will never meet (practically rational from the social point of view). If what is rational from the social point of view is detrimental or indifferent to the non-moral good of a particular person, then would this person still want to do what is morally right?

There is no mystery as to why some people attempt to conform their beliefs to epistemic rationality by striving to form beliefs which are logically coherent, and derived from inductively strong, or deductively sound reasoning. For beliefs which must be reached through reasoning, these strategies are the most reliable way to form true beliefs. And people who reason in this way usually have the goal of forming true beliefs. But will acting in ways which are rational from a social point of view have the same inherent attractiveness as epistemic rationality? As today's global civilization advances, it seems to matter more and more to people that their actions be justifiable from an impartial

perspective rather than exclusively from their partial perspective. If people do value an impartial justification, then Railton’s moral realism should have the required inherent attractiveness, as it provides people an impartial rationality from which to criticize their own and others’ actions and motives.

Although Railton’s moral realism may fail to motivate moral skeptics or those who are indifferent to being rational from a social point of view, this shortcoming can be expected of any moral theory. However, even those who neglect morality can still be held to its standards. Just as those who are indifferent to the law are not thereby excused from its authority, so too those who neglect morality can still be rightly criticized for their moral failings.

1.10 Conclusion

This chapter began with *prima facie* arguments for a fact/value distinction. Railton’s theory provides alternatives to the conclusions of each argument: values can indeed be a species of fact. I explained Railton’s theory of non-moral value, which explains which facts actually constitute non-moral value. Non-moral value, the good for a person, A, can be reduced to facts about A and the world. Specifically the facts about A’s constitution and environment that make it the case that when idealized, A+, would want A to want some things rather than others. The idealization consists in a complete and vivid understanding of all relevant information and perfect instrumental rationality. The good for a person is what is in that person’s objective interest.

This understanding of non-moral value provides us with explanatory power; we use it to explain further facts about the world. Non-moral value allows us to explain why some things that people do turn out better for them than other things they do. It also explains

---

what Railton calls the wants/interests mechanism; people’s subjective interests often evolve with experience to more accurately reflect their objective interests.

With non-moral value in hand, Railton then provides an account of moral value. Each person has objective interests. Morality concerns the interests of multiple people. Morality is impartial to interests of particular individuals, and instead treats all individuals’ interests equally (he calls this impartial perspective the social point of view). Morality is a type of practical rationality, one whose goal is the maximum achievement of the objective interests of all individuals whose interests are at stake when viewed impartially from the social point of view. Thus, moral value is reached through a two-stage reduction. Non-moral value can be reduced to facts about people and their environments. Moral value is then the efficient impartial achievement of the non-moral value of all the individuals concerned.

Moral value also has explanatory power. It allows us to explain why the interactions of groups of individuals sometimes turn out poorly; their actions were not rational from the social point of view. It also has the potential to explain occasional general trends towards the moral rationality in the behavior of groups of people. This trend may be caused by their negative experiences leading them to behaviors that more appropriately mirror rationality from the social point of view.

Thus Railton’s theory of moral value is both reductive and naturalistic. Moral values are reducible through these two stages to natural facts about human beings and the world. It is realistic in that human beings causally interact with the reality of these values.
CHAPTER 2: DEREK PARFIT

Derek Parfit holds a metaethical position of value realism that is contrary to Railton’s value realism. For Parfit, practical rationality is substantive rather than instrumental.\textsuperscript{41} Therefore, the objective facts of a situation are sufficient to provide rational agents with reasons to act. The reasons provided by substantive rationality are what he calls \textit{irreducibly normative reasons}. Irreducibly normative reasons are non-natural in that they do not have ontological implications.\textsuperscript{42} The only way to explain irreducibly normative reasons is through similar terms such as saying that one “ought,” or “should,” or “must” do what these reasons prescribe. This view of substantive rationality leads him to criticize views like Railton’s. Parfit calls his objective non-naturalist view Rationalism.\textsuperscript{43} I will begin by explaining his epistemological and metaphysical justification for his view, in order to dispel some of the view’s prima facie mysteriousness. I will then examine some of his arguments against subjectivism and naturalism.

\begin{itemize}
\item \textsuperscript{42} Parfit, Derek, and Samuel Scheffler (ed). \textit{On What Matters: Volume Two}. Oxford: Oxford Univ Pr, 2011. Henceforth, “On What Matters: Two”. p. 749. Ontological implications are ordinarily said to be had by \textit{anything} that exists. Parfit instead uses existence in the ontological sense to refer to both the ordinary existence of matter in space and time, and perhaps a similar existence of abstract objects such as Plato’s forms outside space and time. However, Parfit takes a deflationary view of the ontological existence of many abstract objects and argues that we ought to remain agnostic as to whether abstract objects exist in this ontological sense outside space and time. Instead we need only admit that due to the role they play in our thoughts, abstract objects have some sort existence and we need not be committed to existence an ontological sense (Parfit, “On What Matters: Two,” pp.475-480).
\item \textsuperscript{43} Parfit, “On What Matters: Two,” p.21.
\end{itemize}
2.1 Parfit’s Metaphysical and Epistemological Justification

The traditional problem for other versions of Rationalism has been their inability to provide plausible metaphysical and epistemological accounts of moral properties. If moral properties are non-natural, they are by definition things that cannot interact with the natural world. If so, how could we ever know that they exist, or know facts about which moral propositions are true? Parfit attempts to succeed where other Rationalist theories have failed. I will examine his metaphysical and epistemological accounts in turn.

2.1.1 Metaphysical Justification

Parfit begins the discussion of the metaphysics of his view by noting that his belief that things matter is his belief that there are irreducibly normative truths. In order for things to matter, there must be irreducibly normative truths, and if such truths exist, then that is what it is for things matter. He confesses that, since he greatly desires that things matter, he is not performing his philosophical investigations from a position of dispassionate disinterest.

He then responds to J.L. Mackie’s objection that if objective values exist, then they are metaphysically queer, i.e., problematically strange, too strange to postulate. Parfit argues that there are indeed abstract entities similar to his non-natural normative truths that we are all forced to admit exist, in at least some sense of the word ‘exist’. He gives an example of such entities: possible events. These events are compatible with causal determinism, and are hypothetical. After having done P, which results in Q, we can look back and say truly that if I had done R, then S would have resulted instead. Few people

would be willing to claim that there are no possibilities other than the events which actually occur, and in doing so we are admitting that possibilities have some sort of existence. If there were no possibilities other than the ones that are actualized, then we would have no reason to plan for the future, nor to regret past mistakes. This concession is too high a price to pay to avoid possibilities. If one admits that there are possibilities, then one has admitted that possibilities exist.

Parfit then moves on to mathematical entities such as numbers. Platonists argue that numbers exist in a realm outside of space and time, and nominalists argue that these things do not exist in any respect. Parfit instead argues that since it is not clear what it would mean to exist outside of space and time, and such a location is too ill-defined to understand, we should not claim that numbers do not so exist. He further argues that this question can be sidestepped completely. On a view that he calls Non-Metaphysical Cognitivism, we can know truths about some things, regardless of whether they exist in any ontological sense.

Nothing could be truer than the truths that 2 is greater than 1, that 2+2 = 4, and that there are prime numbers greater than 100. Not even God could make such claims false. For such claims to be true, there must be some sense in which there are numbers, or in which numbers exist. But in deciding which mathematical claims are true, we don’t need to answer the question of whether numbers really exist in an ontological sense, though not in space and time.

---

So for Parfit, the fact that we can know truths about abstract entities is enough of an existence as need matter to us. These same considerations apply to logical inferences and irreducibly normative truths. So the question is too murky to decide whether abstract entities exist in an ontological sense. Nevertheless, the fact that we can know truths about abstract entities proves that they exist at least in some non-ontological sense.

Parfit provides an example of a hypothetical universe in which nothing ever existed, to try to better explain this conception of non-ontological existence.

If nothing had ever existed in any ontological sense, there would not have been any stars or atoms, nor would there have been space, or time, or God. But it would have been true that nothing ever existed. As we can also claim, there would have been the truth that nothing existed in the ontological sense. This truth would have existed in a different non-ontological sense.51

Thus things can exist in more ways than one. In the ontological sense of exist, there are things like concrete objects which exist in space and time, and perhaps God outside of space and time. In the non-ontological sense, there exist abstract entities, including truths about concrete ontological entities.

Even if abstract objects exist in some non-ontological and thus non-natural sense, why should we believe that irreducibly normative truths are among them? Having shown that it is reasonable to postulate non-natural properties, Parfit next attempts to show why we should believe that some of these are normative truths. He attempts to do so by showing that we have reasons to believe that such truths exist, and so we ought to believe in them, i.e., we have irreducibly normative reasons to believe in them.

2.1.2 Epistemological Justification

Parfit begins his epistemological justification by considering a possible naturalist objection to non-naturalism, which he calls the Causal Objection.\textsuperscript{52} This objection argues that since non-natural entities like irreducibly normative reasons exist only in the non-ontological sense, perception of such entities would require some bizarre sensory organ. Since we have no reason to believe that we have such a perceptual ability, we should believe that even \textit{if} there were non-natural properties, we would never have any way to know about them.

Parfit answers this objection by rejecting the notion that non-natural properties are contingent entities that need to be perceived as existing out in the world.\textsuperscript{53} Instead, we are able to form true beliefs about non-natural properties because truths about non-natural properties are necessary truths. We form our true beliefs about them analogously to the way we form true logical and mathematical beliefs. They may or may not be “analytic,” by which Parfit means true by definition.\textsuperscript{54} However, these beliefs are formed by examining the content of the beliefs themselves. In general, in all domains, we form true beliefs either by recognizing their intrinsic credibility (by which he means intuitive appeal, or that something \textit{seems} to be true), or by responding to the reasons we have to hold a belief. Likewise, true beliefs about non-natural properties can likewise be formed in either of these ways.\textsuperscript{55}

A further objection which follows from the Causal objection, Parfit calls the Massive Coincidence Objection.\textsuperscript{56} This objection runs, since we cannot interact causally with non-natural abstract entities like numbers or irreducibly normative reasons, it would be a massive coincidence if our beliefs about them were true. We should not assume such an implausible coincidence to be the case. Therefore, non-naturalism must explain how we causally interact with them to form our true beliefs about numbers and normative truths.

Parfit again replies that although we are indeed not causally related to non-natural entities, we can nonetheless respond to them in a different way.\textsuperscript{57} We do this by responding to reasons. He gives an example of a computer as something which we have made and which can correctly calculate mathematical truths without being causally affected by numbers. Likewise, human brains may also simply respond to reasons without being causally affected by them.

When some fact has the property of being or giving us a reason, we cannot be causally affected by this normative property. But we respond to such properties in other ways. We respond to reasons when we are aware of the facts that give us reasons, and this awareness leads us to believe, or want, or do what these facts give us reasons to believe, or want, or do. Therefore, the dilemma posed by the naturalists is a false one. We need not causally interact with non-natural properties in order to form true beliefs about them. Instead of interacting causally with reasons we interact \textit{rationally} with them. We respond to the reasons we have to believe one thing or another.

In fact the objection can more plausibly be run in the other direction, and Parfit calls this the Reverse Coincidence Argument.\textsuperscript{58} We have reasons to believe that we form true mathematical beliefs, due to the way in which science has utilized math to make accurate predictions about the world.

Unlike other animals, we form true beliefs about what we cannot see, hear, touch or smell, such as beliefs about possibilities and the further future. We have this ability, we can claim, because we can respond to epistemic reasons. If we were not responding to epistemic reasons, our ability to form so many true beliefs would involve a highly implausible coincidence.\textsuperscript{59}

Skepticism about the fact that human beings form true beliefs by in some way responding to non-natural properties is far less justified than the mere belief that humans are able to respond to reasons. Therefore, more massive coincidence is required by naturalism than by non-naturalism.

But why think that responsiveness to reasons is rational rather than causal? On the relevant distinction between natural and non-natural properties, properties are natural if they are studied by the natural or social sciences.\textsuperscript{60} Parfit provides validity as an example of a non-natural abstract property.

When we call some argument valid, we mean that, if this argument’s premises are true, this argument’s conclusion must be true. The necessity of these truths is not part of the causal fabric of the world. When Metaphysical Naturalists reject Platonism, these people rightly deny that we could be causally affected by abstract entities or properties. Validity is

one such property. Since this explanation of our abilities must appeal to
the fact that our reasoning is valid, which is not a natural fact, this
explanation is not wholly naturalistic.\(^{61}\)

The concession that the abstract property of validity exists shows that irreducibly
normative truths exist. We are able to respond to the validity of arguments. This response
is a non-natural, non-causal response, through which we are able to respond to irreducibly
normative reasons.\(^{62}\)

Parfit also draws the distinction between normative and non-normative senses of
probability.\(^{63}\) In the the non-normative sense, probability is either a statistical correlation or
a logical necessity (Parfit calls this the “alethic” sense of probability). In the normative
sense, if a statement is probably or certainly true in the non-normative sense, then one
\textit{ought} to believe it.

When we know that certain facts have the alethic property of implying that
P is true, that makes these facts have a different normative property of
giving us a decisive reason to believe P.\(^{64}\)

Thus, when a belief, P, is likely to be true, then this fact has two separate properties. First,
it has the non-normative property that P is likely to be true. Second, it has the additional
normative property of providing one with an irreducibly normative reason to believe that
P.\(^{65}\) The distinction between these two properties shows that Parfit’s irreducibly normative
reasons exist, at least for epistemology, by virtue of the fact that when one thing makes
another likely to be true, this fact provides one with a reason to believe the other. For

example, the fact that it is raining outside makes it likely that the ground outside is wet. Thus, the fact that it is raining outside also provides one with a reason to believe that the ground outside is wet.

Even supposing that this distinction shows that we have irreducibly normative epistemic reasons to believe, why should we believe that we have irreducibly normative practical reasons to act? Here Parfit argues that the intuitive plausibility of our normative practical reasons are again enough to justify our belief in them. Analogous to the fact that by knowing P, and that given P, Q is likely to be true, we are provided with irreducibly normative reasons to believe Q, so too the fact that S will cause us great suffering provides us with irreducibly normative reasons to avoid S. This is a fact that is simply intuitively obvious to us as human beings.

Though we could not possibly have empirical evidence for these beliefs, we also could not have much evidence against these beliefs. When we ask whether we can have practical and moral reasons, nothing is relevant except our normative intuitions. If it seems to us to be clearly true that we can have such reasons, and we seem to have no strong reason to believe that we can’t have such reasons, we can justifiably believe that we can have such reasons.

Because reasons are abstract non-natural properties that exist in a non-ontological sense and in no way causally interact with the concrete spatio-temporal reality of our world, empirical evidence is irrelevant to the credibility of our beliefs in reasons. However, given that we are relying solely on intuitions, Parfit needs to show that there is substantial agreement among our intuitions. If everyone claims to have intuitions telling them to do

---

different and conflicting things, the plausibility of them actually representing reasons would be questionable. This is the problem that we considered in section 1.8. and called the argument from relativity.

To answer this objection Parfit defends what he calls the Convergence Claim (CC). The CC is the assertion that if certain confusions are accounted for, then the normative intuitions of the majority of humanity converge rather than diverge.

If everyone knew all of the relevant non-normative facts, used the same normative concepts, understood and carefully reflected on the relevant arguments, and was not affected by any distorting influence, we and others would have similar normative beliefs.68

This claim is basically the denial of the argument from relativity, so the strength of this rebuttal by Parfit will depend on his justification for such a claim. Parfit notes that such justification will be difficult and complex and gives certain qualifications on what would be sufficient justification for the claim.69 First, he notes that this is an empirical claim, so we are considering how people actually act under certain conditions. Second, he notes that given the fallibility of human beings, it is only necessary that almost everyone agree (some will always make mistakes), and also that this agreement need merely be adequate rather than total agreement. Third, it will be difficult to define ideal conditions and distorting influence, and we must be careful not to define them so as to beg the question for or against the CC.

Parfit defends the CC by considering some examples of seeming intuitive disagreements, and then explaining them away. The first is that the nature of agony

---

makes it the case that we all have a reason to avoid being in agony.\textsuperscript{70} He claims that the only people who would doubt this claim are those who do not believe that such reasons could possibly exist. However, we should discount these views because including this disagreement would skew the investigation against the CC. Instead it is enough that such people would agree that whatever analogous concepts are used in place of reasons in their framework would agree that the nature of agony leads us to recommend against it.

Next Parfit considers cases of public policies.\textsuperscript{71} There is vast disagreement as to which policies ought to be instituted by any particular government. These policies have great moral consequence on the lives of the people to which they apply. In these cases, argues Parfit, the disagreement is probably due to the relevant contingent empirical facts of the situation. Different people imagine that each policy will actually result in different amounts of suffering than their opponents imagine, or hold different views of human nature, or have different beliefs about God’s commands. In ideal conditions, everyone would agree on these non-moral empirical facts of the situation and their disagreements about which policies we have reason to enforce would disappear.

In other cases, people disagree due to distorting influences, such as the circumstances of their particular situation.\textsuperscript{72} For example, when considering the duties owed by the rich to the poor, people’s intuitions may be unduly influenced by whether \textit{they} are in fact rich or poor. Under ideal conditions for reflection people would not consider whether a moral decision will benefit them rather than others and so this disagreement

would disappear. Similarly some people may just be using moral terms differently, and if they used clear definitions many disagreements would disappear.\textsuperscript{73}

Other moral disagreements may simply stem from the ethical theory to which people subscribe.\textsuperscript{74} People with differing ethical theories most often agree on which acts are right or wrong, but disagree on the right-makers or wrong-makers of such actions. For instance Kantians and utilitarians both agree that killing innocent people is generally wrong. However, Kantians believe that murder is wrong because it violates the categorical imperative, and utilitarians believe that murder is wrong because it does not maximize happiness. Parfit argues that we can likewise disregard these disagreements because plausible competing ethical theories tend to lead to convergence, all climbing the same moral mountain from different sides.\textsuperscript{75}

Other disagreements may stem from the fact that the particular problem being considered is a borderline case.\textsuperscript{76} Again, for example, we almost all agree that it is generally wrong to kill innocent human beings. However, many disagree about who or what should be considered human beings. Fetuses are an example of borderline humans. Others may disagree about what constitutes killing rather than allowing to die. Such disagreements about borderline cases should not count against Parfit's claim that our intuitions support the claim that we have irreducibly normative reasons not to kill innocent human beings.

\textsuperscript{74} Ibid.
Other disagreements may stem from the mistaken belief that right and wrong are binary concepts, rather than scalar concepts.\textsuperscript{77} If people believe that right or wrong is all or nothing, rather than rightness and wrongness coming in degrees, then they will tend to err towards the extremes. For example, if you are morally obligated to give to the poor, and there are always more people whose lives you could save by giving just a little bit more, then you must be obligated to give away pretty much everything you have. Since doing so seems morally optional rather than required, we might then conclude that we have no obligation to give to the poor. If instead everyone agreed that rightness and wrongness come in degrees, then everyone would just agree that in general the more that one gives to charity, the more right and the less wrong one would be doing.\textsuperscript{78} Similarly, the older and more like a human being a fetus becomes, the more wrong killing such a being would be. Since these disagreements would dissolve if everyone agreed that rightness and wrongness come in degrees, they should not count against the CC.

Lastly, argues Parfit, many moral disagreements occur because people assume that all normative truths must be precise and determinate.\textsuperscript{79} In fact, things are a lot messier. Not everything can be judged by a single numerical value along a single scale. Since there may incommensurable values at play in some cases, there may not be any exact truth as to which option is better. However, the mere fact that in some cases there is no fact of the matter as to which of two options is better does not entail that there are not many other cases in which there is a fact of the matter as to which of two options is better.

\textsuperscript{78} Here the idea meant here by ‘in general’ is that most people in the affluent world could give a lot more to charity than they do, and the more they gave without harming themselves or their loved ones, the more right their actions would be. The idea should not be taken to mean that people who give to the point of losing their ability to provide for themselves, or starving themselves to death would thereby be doing the right thing.
The mere presence of gray areas does not entail that there are not also areas in which the moral truth is clearer. Since these disagreements are over the nature of normative truths, they would probably be alleviated if everyone understood that in some cases the truth is imprecise and indeterminate. Thus, these disagreements should not count against the CC.

Given all these ways in which intuitive moral disagreements are due to the lack of ideal conditions for moral reflection, we are justified in believing the CC. Since Parfit has argued that reasons are abstract non-natural properties that exist in a non-ontological sense, and in no way causally interact with the concrete spatio-temporal reality of our world, empirical evidence is not necessary to the credibility of our beliefs in such reasons. Therefore, all that is relevant to the truth of our beliefs about irreducibly normative truths is our intuitions about the matter. Since the CC shows that our intuitive moral beliefs are probably sufficiently similar, this provides us with the justification that we need to believe that there are irreducibly normative truths. Therefore, we have irreducibly normative reasons to believe that there are irreducibly normative truths.

### 2.2 The Smoker and Agony

In this section I will examine an example in which Parfit is a smoker and an example in which he has subjective reasons to cause himself to be in agony for its own sake. These arguments are designed to appeal to our intuitions about what we have reasons to do. They are meant to suggest that the reasons provided by subjectivist theories\(^80\) like Railton’s are unappealing, while the reasons provided by Parfit’s Rationalism are appealing.

---

\(^{80}\) By subjectivist theories Parfit means all theories in which people’s reasons are derived from their desires or idealized desires rather than irreducibly normative reasons.
Many of Parfit’s arguments involve only what a particular hypothetical person has reason to do in regards to his own interests, which is what Railton refers to as non-moral value. In one such example we suppose that he is a smoker:

We can suppose that, unless I stop smoking, I shall die much younger, losing many years of happy life. According to all plausible objective theories, this fact gives me a decisive reason to want and to try to stop smoking.\(^8\)

Parfit takes it as obviously true that anyone in this situation *is* provided with such a decisive reason. Rationalism accounts for this reason. Thus, it is a mark against subjectivism that it cannot universally claim: the fact that smoking causes one to lose many years of happy life provides one with a decisive reason to quit. Remember for Railton, a person’s decisive reasons are provided by that person’s objective interests. A person, A, has decisive reasons to quit only if A+ (fully informed and instrumentally rational A) would want A to want to quit smoking. As Parfit points out, it is theoretically possible that A may be so constituted that continuing to smoke is more important than the likelihood of enjoying many extra years of happy life.

Parfit’s view does have some intuitive appeal. It does seem that many years of happy life are clearly something that most of us would judge to be more worthwhile than the pleasures of smoking for *anyone*, regardless of how much they like smoking. When we consider people who continue to smoke despite knowing the risks, we assume that they either consider themselves to be special exemptions to the rule that smoking usually shortens peoples lives or else they are too addicted to do what they know is good for themselves. It may in fact be true that if for *any* person, A, (a) it is true that smoking will

---

take years off of A’s life, and (b) those lost years would be happy and enjoyable, and (c) A is able to know (a) and (b) and vividly consider them, then A would want to quit smoking. If for any person, (a), (b), and (c) are all true, then Railton’s theory would agree: the fact that smoking causes one to lose many years of happy life does provide anyone with a decisive reason to quit. All Railton’s view rules out is that this decisive reason is necessarily provided to everyone \textit{a priori}. If it is true that for any person A, that A+ would want A to quit smoking, then it is the case that everyone would have a decisive reason to quit. However, this is an \textit{a posteriori} fact which we cannot claim to know beforehand. For Parfit, (a) and (b) are sufficient to provide anyone with a decisive reason to quit, regardless of what the features constituting the person determine to be one’s objective interest.

Another way in which Parfit’s example may fail is that it includes the possibly value-laden term happy. If by “many years of happy life,” Parfit means “many years which everyone would prefer to have rather than continuing smoking,” then the fact that smoking causes one to lose many years of happy life \textit{does again} provide one a decisive reason to quit. And this decisive reason does not require substantive rationality. The hypothetical imperative, “If I would prefer many years of happy life to smoking, then I have a decisive reason to quit smoking,” trivially yields the consequent, given that the antecedent it true by hypothesis.

In a similar vein, Parfit provides an argument against subjectivism that concerns agony, illustrated by this case:

I want to \textit{have} some future period of agony. I am not a masochist, who wants this pain as a means to sexual pleasure. Nor am I a repentant sinner, who wants this pain as deserved punishment for my sins. Nor do I
have any other present desire or aim that would be fulfilled by my future
agon. I want this agony as an end, or for its own sake. I have no other
present desire or aim whose fulfilment would be prevented by this agony,
or by my having my desire to have this agony. After ideal deliberation, I
decide to cause myself to have this future agony, if I can.82

In this case, Parfit explicitly stipulates that A+ determines A’s good to consist in what we
all intuitively see not to be in A’s good. If a subjectivist account of reasons cannot rule out
such a case, then it must be false.83 Again Parfit’s objective account has intuitive appeal.
Parfit plausibly claims that we could never desire agony for its own sake. Thus, it is a
small step to the conclusion that substantive rationality provides everyone with irreducibly
normative reasons to avoid agony.

Again, Railton’s theory does not rule out, in principle, the possibility that some
people may be so constituted that things are in fact be good for them which we find
counter-intuitive. But his view does not entail that he lacks a plausible account of the good
for a person. It simply entails that if we postulate a person with a constitution that is
extremely different from ours, such a person’s desires may seem puzzling.

Since in this example A+ actually determines that it is best for A to experience
agon, we can rule out the possibility that “agon” means “something that no being could
desire to experience.” Agony may instead simply mean excruciating pain. There may, in
fact, be no person on Earth, A, whose idealized self A+, would want A to be in agony. A+’s
perfect knowledge of what the experience of agony entails would prevent A+ from ever
recommending it. However, it should not count against Railton’s theory that it allows that if

there were a person A, whose idealized self A+, would want A to be in agony, then being in agony would actually be good for A.

In fact it should count as a merit of Railton’s account that the good for A depends on the constitution of A, rather than what people of extremely different constitutions hold to be good for A. Suppose for example there were a human being with a strange genetic abnormality which made it so that arsenic, rather than poisoning and killing him, would make him stronger, smarter, and healthier. It seems to be uncontroversial to say that this person has reasons to eat arsenic, and that eating arsenic would be good for him. We might be tempted to say that everyone has a reason not to eat quadruple the lethal dose of arsenic. Yet as the arsenic man example shows, we cannot know this a priori to be the case because there could be a person for whom arsenic is beneficial. Yet this is exactly analogous to the Parfit agony example above. Parfit imagines a person for whom agony is good, and is then surprised when agony is good for him. We cannot make universal claims about the badness of agony while at the same time postulating individuals with constitutions vastly different than any we have encountered.

Anticipating objections to this argument, Parfit argues that it is not enough that the idealized self, A+, of every actual agent A, would want A to not be in agony. In other words, it is not enough to claim that subjectivism agrees that everyone does have reasons to avoid agony, if it is possible to imagine beings that have subjective reasons to be in agony. If hypothetical beings would have reason to want to be in agony, then subjectivism must be false.\(^\text{84}\) To support this demand, Parfit argues that if the mere fact that A+ wanted A to want something were to provide A with reasons to want that thing, then we might have reasons to do all sort of implausible things such as being in agony, or wasting our

lives. These claims, Parfit asserts, are clearly false. Thus, the demand that decisive reasons be provided to everyone a priori is justified. A subjectivist account must provide irreducibly normative reasons for all rational agents regardless of their contingent constitutions. However, Parfit is wrong, and such a demand is not justified. Assuming that reasons cannot be derived from actual or possible states of a subject is to beg the question against the subjectivist who claims that reasons are desire-based, and reasons derive their normativity from the desires of idealized agents.

Thus, the objection that subjectivism fails because a person could never have a decisive reason to desire agony is unsuccessful. The objection holds only if we are either imagining actual people whose experience of agony is similar to ours or if we define “agony” as “something that no being could desire to experience.” But if either of these options is the case, then there would be no A+ who would want A to want to be in agony. Given the nature of agony, it seems highly likely that no actual people could come to desire agony for its own sake. Despite this contingent fact, the mere possibility that there might exist a person who when idealized, did desire agony for its own sake, does not entail that agony could not be good for such a strange person. Therefore, the smoker and the agony arguments do not prove as much as Parfit may have hoped.

2.3 Argument from the Source of Reasons

Parfit provides another argument that subjectivism is incapable of providing us with reasons. This argument could be called the argument from the source of reasons. The argument from the source of reasons says that any instrumental reason must be derived

either from another instrumental reason or from an intrinsic reason. Because subjectivism rests all chains of reasons on idealized desires, subjectivism cannot be a source of reasons.

According to these people [subjectivists], instrumental reasons get their force, not from some telic reason, but from some telic desire or aim. We can have desire-based reasons to have some desire, and we can have long chains of instrumental desire-based reasons and desires. But at the beginning of any of these chains, as we have seen, there must always be some desire or aim that we have no such reason to have.\(^88\)

On a subjectivist view such as Railton’s, our intrinsic reasons are derived from the desires of our idealized selves. Since our idealized selves have no reason to desire what they do, our source of reasons is without a foundation, and subjectivism fails.

But are the desires of our idealized selves really, “some desire or aim that we have no such reason to have?” By “no such reason”, Parfit simply means no *irreducible* *normative reason*. But of course one does not have irreducibly normative reasons to have desires because subjectivist theories do not claim to provide irreducibly normative reasons. To demand an irreducibly normative reason is to beg the question against the subjectivist. In particular Railton’s subjectivist theory reduces the good for a person, A, to what an idealized version of that person, A+, would want A to want. A’s reasons are then provided by these objectified interests of A.

In fact the desires of our idealized selves do provide us with reasons in a reducible sense of the concept ‘reason’. The fact that all chains of reasoning terminate in the desires

---

of an idealized agent does not imply that these desires are arbitrary or whimsical, or in any way problematic. In fact, similar to Parfit’s theory, these desires are grounded in part in the objects of desire. On Parfit’s Rationalism, it is the fact that broccoli is healthy that gives me a reason to eat broccoli, rather than my idealized self’s desire that I eat broccoli that gives me a reason to eat broccoli. But why would my idealized self want me to want broccoli? Idealized agents are forced to take into account the fact that broccoli is healthy. However, this fact alone does not entail that my idealized self will desire it. If my existing repulsion of the taste and texture of broccoli makes the consumption of broccoli unpleasant, and the effacing of this repulsion would require many hours of painful therapy, and my idealized self knows that my dietary needs can be readily met through the plethora of other vegetables that are available and that I enjoy, then my idealized self may not desire me to seek broccoli. The desires of idealized agents take into account all of the objective as well as the subjective features of the situation. To look only at the objective features is to miss half of the picture.

On a subjectivist view such as Railton’s the desires of an idealized agent are grounded in the facts of the world, including facts about the subject for whom things are good or bad, and the features that constitute the subject, which make some things in the world better and other things worse for the subject. It is the relationship between the subject and the world that grounds the desires of the idealized agent which in turn gives his desires their normative force. Railton’s theory is a plausible and powerful explanation of why we have a reason to eat broccoli. I think that Parfit’s objection is simply underestimating the resources that a sophisticated subjective ideal agent account of the good like Railton’s has to consider and include the objective features of the objects in the world.
The main disagreement between Parfit and Railton is not about what we have reasons to do. Both agree that we have reasons to do what is in our best interests, which for most people means that we sometimes ought to eat broccoli. The disagreement is about the metaphysical account of these reasons. Parfit thinks that reasons must by their nature be non-natural. Railton instead provides a plausible naturalistic account of reasons. On Railton’s account it is indeed true that our reasons, if we have them, to eat broccoli derive from subjective desires. However these are not “some desire[s] or aim[s] that we have no such reason to have.” A+’s subjective desire does has an objective basis in the facts of the world that make it the case that A+ does indeed want A to want to eat broccoli. This basis, with both subjective and objective elements, seems to me to be a plausible and sufficient source of reasons.

2.4 Hard Naturalism

Another argument of Parfit’s which is applicable to Railton’s moral realism is his argument against what he calls Hard Naturalism. Hard naturalism is any naturalism in which morality is completely reducible to natural properties, and normative statements can be restated completely in descriptive terms. Railton’s theory meets this definition of hard naturalism. For Railton the non-moral good for each person, A, is what A+ would want A to want. Moral good is then the what is rational from the social point of view, i.e., the rational maximization of the non-moral good of everyone concerned when considered from an impartial perspective. These values are constituted entirely in descriptive terms. Thus, Railtons theory is a hard naturalist theory.

As with the agony argument, Parfit again attempts to motivate our intuitions of the implausibility of hard naturalism through an example where it leads to seemingly
implausible recommendations (this time that a woman with anorexia ought to starve herself to death). However, Parfit agrees that these are reasonable recommendations, given the naturalist’s assumptions. Parfit instead concludes that we should not be bothered by these recommendations of hard naturalism, because naturalism is ultimately a trivial rather than a substantive theory.

The main target of this argument is the full-informational account of the good by Richard Brandt, which is a direct forerunner of Railton’s moral realism. In Brandt’s account what one ought to do is to fulfill one’s rational desires, or what would fulfill one’s desires, were one fully rational.\textsuperscript{90} However, Brandt has a precise technical definition of “rational desires,” i.e., one’s desires are rational if and only if one has fully contemplated all available information, and applied logic, to the point where further contemplation of the information would not alter one’s desires in any way. This process of becoming rational Brandt calls “cognitive psychotherapy.” This account of a person’s good is similar to Railton’s view except that on Brandt’s view the idealized agent is limited to actually available information, rather than having the total empirical knowledge (which includes full accurate predictive knowledge of all consequences) that Railton’s view comprises. Since Parfit ultimately rejects the view as trivial due to its hard naturalism rather than the amount of information available to individuals, his objection should apply \textit{mutatis mutandis} to Railton’s.

Parfit begins with his standard move. He notes that it still conceivable that one might be fully rational and desire something which we find to be horrible. In this case a woman with \textit{anorexia nervosa} might still prefer to starve herself to death even after fully considering her options. Finding this intuitively appalling he then asks, “Why ought I to do

what is in [Brandt’s] sense rational?" Parfit notes that Brandt thinks that such a question is not a damaging one as it could be asked of any theory of the good. For example, given hedonism, one can always ask why I should do what will give me pleasure. The hedonist responds because that is what my theory holds to be the good. However, Parfit notes that the question does not make sense against his own theory. Since “why” asks for reasons, it would ask, “What reasons have I to want only those things that I have reasons to want?” This question is not an open question.

However, although Parfit fails to mention it, in addition to noting that any theory must face such questions, Brandt does also provide reasons (information the knowledge of which will incline one to agree with Brandt, so reasons in Brandt’s sense) why one ought to do what is in his sense rational. Brandt argues that everyone agrees that we should use facts and logic in decision making, and it is arbitrary to limit consideration to just some of the available facts or some of the probable outcomes. As a matter of fact, people are unhappy when their decisions lead to outcomes they do not enjoy, and these outcomes were brought about solely because they failed to adequately reflect on the available information. People do not want to make choices that they will later regret due to this type of negligence. Doing what is rational is to avoid this mistake. Therefore we ought to do what is rational. Brandt also makes a second argument. Every person in fact has desires. Doing what is rational, i.e., considering all of one’s options as fully as possible, is most likely to lead to the fullest satisfaction of one’s desires. You do want the fullest satisfaction of your desires. Therefore, you ought to do what is rational.

---

Parfit argues that we should not actually be troubled by the outcome of Brandt’s theory in Parfit’s imagined scenario--namely, that the woman ought to starve herself to death--because Brandt’s theory is actually trivial. In saying that starving to death would be best for the anorexic woman and what she ought to do, “We would mean only that, in starving herself to death, this woman would be doing what, even after cognitive psychotherapy, she would want most.” Parfit defines a substantive normative claims as one that both

(a) states that something has some normative property,

and

(b) is significant, by being a claim with which we might disagree, or which might be informative, by telling us something that we didn’t already know.

Since Brandt’s theory provides a reduction of what we ought to do, to what we would want to do if in his sense rational, Parfit thinks the account of reasons no longer involves a normative property. However, this is tantamount to begging the question against the reductionist who claims that normative properties are reducible to natural ones. By assuming that Brandt’s theory cannot have normative properties simply because it does not invoke irreducibly normative reasons, is to rule out all other conceptions of normativity by definition. Brandt’s theory is substantive rather than trivial because it argues that you do want to do what is rational, and thus provides recommending force. This is significant because it is both something about which we can disagree, and which would be informative to someone who has not considered that we have reasons to want what we would want if in Brandt’s sense rational. For instance starving herself to death is, ex hypothesi s (in this extraordinarily far-fetched scenario), what the woman would want if

rational. Since, Brandt’s theory states that she therefore has reasons to do this counterintuitive action, it is significant; without Brandt’s theory we would not think that she has reasons to do this action. But, given the way human beings are constituted, doing what is rational best achieves what they really want in life. If one wants to achieve what one really wants in life, then one will do what is rational. One does want to achieve what one really wants in life. Therefore, one ought to do what is rational. Therefore, the woman ought to starve herself to death. The force of the ought is reducible to a hypothetical imperative, the antecedent of which, given the facts of the natural world, are true.

This may seem to be an unacceptable result for any theory of the good for a person. However, keep in mind that for any actual human being with anorexia, starving oneself to death would not be what one wanted after cognitive psychotherapy. People with anorexia do not want to starve themselves to death. They want to be thin. They have a distorted self-image in which they mistakenly believe themselves to be fat despite actually being thin. If they were to fully and vividly contemplate the facts of their situation, they would come to see that their desire was grounded in their false belief that they were fat. This desire would be effaced as they came to realize and appreciate the truth. In practice in the actual world, undergoing cognitive psychotherapy may be very difficult, and many people may be unable to contemplate the facts to the point where the facts have all of the effects that they would if considered fully. This harsh practical reality does not preclude it from being a plausible theoretical account of the good. Thus, plausible subjective metaethical theories of the good for a person need not yield Parfit’s seemingly absurd conclusions. In addition, both Brandt’s and Railton’s theories are substantive rather than trivial theories, because in addition to providing a reduction of non-moral value, they argue that we have reasons to want these values to be realized.
Parfit attempts to further illustrate the triviality of hard naturalism through an example in which he is critical of Frank Jackson.\textsuperscript{95} Again, since Railton’s theory is also one of hard naturalism, the objection should apply \textit{mutatis mutandis} against his theory. Jackson agrees that if naturalism is true, then normative terms would be, in principle, dispensable. Parfit quotes Jackson that if the metaethical view that hedonism was the moral good turned out to be true, then

‘we should identify rightness with maximizing expected hedonic value’

because this would be what ‘we ought to aim at’.

Parfit notes that if normative terms are dispensable, then this statement could be rephrased as,

it would maximize expected hedonic value to identify rightness with maximizing such value, because this would be what it would maximize such value to aim at.\textsuperscript{96}

The analogous argument raised against Railton’s theory would state,

we should identify rightness with maximizing value from the social point of view because this would be what we ought to aim at.

Becomes,

It would maximize value from the social point of view to identify rightness with maximizing such value, because this would be what it would maximize such value to aim at.

These views, due to their hard naturalism turn out to be trivial rather than substantive theories. Since any occurrence of ‘ought’ or ‘should’ can be replaced with the naturalized reduction of this term, hard naturalism does not actually recommend anything! But here


Parfit is merely playing with words. He is equivocating between moral and pragmatic normative statements. This trivial interpretation is not what Railton, nor Jackson, is saying at all. When Railton says we “should” identify rightness with “maximizing value from a social point of view”, Railton means, “if we want to provide a reforming definition of moral rightness which captures everything important about morality and yet casts moral concepts in realistic, understandable, and naturalistic terms, then we will identify rightness with maximizing value from a social point of view.” Not every use of ‘should’ expresses a moral sense of the term. Some instances are merely pragmatic, and suggesting how to most effectively order our ethical discourse. By “we ought to aim at” maximizing value from the social point of view Railton is not postulating irreducibly normative reasons nor is it a trivial claim. It actually means “If you want to do what is right, then you must maximize value from the social point of view.” This hypothetical imperative provides normativity and information by prescribing what must be done by those who want to do what is right.

Thus, Railton, Brandt, and Jackson’s views are substantive accounts of moral realism. Consider the staple example of a non-moral reduction: water is H₂0. Although we can substitute water for H₂0 that does not mean that we are forced to do so in all areas and instances or our discourse. When asked about H₂0, you can also describe it as water. When asked about water we can provide further information about it by explaining that it is composed of two parts hydrogen and one part oxygen. Hard naturalism may, in principle, be able to reduce value to completely descriptive naturalistic terms. That does not make hard naturalism trivial. Nor does it entail that abandoning traditional evaluative and normative language will be the most practical way to order our discourse on the topic.
2.5 Naturalism and the Normativity Objection

Hard naturalistic theories of value such as Railton’s theories of moral and non-moral value are substantive, in that they provide us with information and reasons. But could Parfit be right and these theories are in some way crucially lacking? It is true that they do not provide us with irreducibly normative reasons. But why must a theory of value provide us with irreducibly normative reasons rather than naturalistically explicable reasons? Are the reasons provided by Railton’s theory all that are required for realistic theories of moral and non-moral value? Parfit thinks that naturalism is incapable of providing the type of normativity moral realism requires, and he calls this the Normativity Objection. It concerns the nature of reasons. This objection is close to the argument from the source of reasons. That argument claims that desires, even idealized desires, do not provide us reasons. The Normativity Objection instead concedes that naturalism may provide us with reasons. However, these reasons are just too different from the type of reasons that a successful theory must possess for naturalism to be viable.

Parfit argues that the difference between irreducibly normative reasons provided by Rationalism and the reasons provided by naturalistic theories is a difference of kind rather than degree. Just like the difference between a color and a shape, the reasons of these respective theories are of different categories.

It could not be a physical or legal fact that $7 \times 8 = 56$, nor could it be a legal or arithmetical fact that galaxies rotate, nor could it be a physical or arithmetical fact that perjury is a crime. It is similarly true, I believe, that when we have decisive reasons to act in some way, or we should or ought
to act in this way, this fact could not be the same as, or consist in, some
natural fact, such as some psychological or causal fact.\textsuperscript{97}

Natural facts \textit{seem} to be just too different from the ideas that Parfit means when he uses
the words 'should' or 'ought,' for the natural fact to be the same kind of things as these
normative concepts. A Naturalist might respond that when providing reductive definitions
of normative concepts, the Naturalist is not claiming to be providing definitions of what
people mean when they use normative terms.\textsuperscript{98} A reductive definition can take one of two
forms. Analytic reductionists claim that their reductive definitions are what people \textit{actually}
mean when they use normative terms. Non-analytic or synthetic Naturalists instead claim
that our ordinary uses of normative concepts are vague and they have no single clear
meaning. By providing reductive definitions, synthetic Naturalists claim to be capturing the
relevant components of the meanings of our normative concepts. We can appeal again to
the analogy of water and $\text{H}_2\text{O}$. People throughout history have known and were able to
recognize and refer to water. However, scientists were able to discover that water had the
underlying molecular structure of $\text{H}_2\text{O}$. These scientists were not concerned by what
laypeople meant when they said water, these scientists were discovering knowledge and
providing us with further information about water.

Parfit rightly objects that the meaning of the word water does affect the discovery
of the fact that water is $\text{H}_2\text{O}$.\textsuperscript{99} It is because of what people meant by water that the
discovery of $\text{H}_2\text{O}$ was a discovery about \textit{water} and not some other or new substance. The
original meaning of water constrained the possibilities of what water could be made. Given
what people meant by water, it could not possibly be the case that the underlying structure

of water was a musical note or a person’s earlobe. The original meaning of water simply eliminates some categories form the list of water’s possible underlying structures. Likewise, claims Parfit, there is an unbridgeable category difference between what he means by ought and what Naturalists want to reduce it to.

Parfit attempts to demonstrate this category difference through an example he calls the ‘Burning Hotel’. In this example, you are on an upper floor of a burning hotel. If you stay in the hotel you will die, but luckily there is a canal next to the hotel into which you can jump from a window, saving your life.

Since your life is worth living, it is clear that
(B) you ought to jump.

This fact, some Naturalists claim, is the same as the fact that
(C) jumping would do most to fulfill your present fully informed desires, or is what, if you deliberated in certain naturalistically describable ways, you would choose to do.

Given the difference between the meanings of claims like (B) and (C), such claims could not, I believe, state the same fact. Suppose that you are in the top storey of your hotel and are terrified of heights. You know that, unless you jump, you will soon be overcome by smoke. You might then believe, and tell yourself, that you have decisive reasons to jump, that you should, ought to, or must jump, and that if you don’t jump you would be making a terrible mistake. If these normative beliefs were true, these truths

---

could not possibly be the same as, or consist in, some merely natural fact, such as the causal and psychological facts stated by (C). ¹⁰¹

Thus, just like the difference between H₂O and musical notes, naturalistic and non-naturalistic reasons are separated by a categorical chasm. The things people mean when they use normative terms constrains the possible entities to which the terms can possibly refer. Reforming naturalistic definitions of these terms fall on the wrong side of this divide.

Parfit’s approach is standard: he attempts to make the Naturalists’ conclusions clash with our intuitions about what is going on in a seemingly obvious situation. However, his example is deceptive in several ways. What we all know about the relations of fire, pain, suffering, and agony provide us with intuitive certainty that in the Burning Hotel example we would have decisive reasons to jump. Parfit asks us to examine our intuitions about the metaphysical nature of these reasons. Doesn’t it seem likely that these reasons are solely objective and provided by substantive rationality and apply equally to everyone? Well yes, given the obviousness of the situation, that does seem to be a plausible. However, a naturalistic theory such as Railton’s provides a plausible account as well.

As stated in Parfit’s (C), Railton’s theory of non-moral value or the good for a person is determined by the desires of an idealized version of that person. A’s good is what the ideally informed and deliberative A+ would want A to want. This account provides an extended sense of rationality which explains why A has reason to do things that are good for A, but that A is not currently disposed to desire or pursue. Due to the natural limitations of actual people, our subjective interests are often skewed. What it is rational for us to pursue given our actual beliefs and desires, often fails to reflect perfectly what is good for us. Thus, Railton’s theory provides us with a naturalistic way to critique our

subjective interests. However, given the obviousness of the situation in the Burning Hotel example, there is no need to invoke this extended sense of rationality; our subjective interests will adequately reflect our objective interests in this case. We can see that jumping will save our lives and prevent excruciating suffering. There is no need to postulate non-natural reasons to explain the fact that we ought to jump. This “ought” can be explained by the completely naturalistic facts that “If you want to live rather than die a horrible death, then you must jump,” and “You do want to live rather than die a horrible death”. These are causal and psychological facts about the world, and if Railton is right, they provide a plausible explanation for the normativity of “You ought to jump.”

We all understand the gravity of the hotel example and the fact that everyone has a decisive reason to jump. We can all see the normative force in the statement “You ought to jump.” Naturalism accounts for the urgency of the situation (you really want to live rather than suffer). Since both the naturalistic and non-naturalistic accounts of normativity agree that everyone has a decisive reason to jump, this case does not support the ruling that naturalism should be ruled out as a metaphysical non-starter. Naturalism is just as plausible of a candidate for the role in normativity analogous to the role of H2O in water.

Due to the obvious decisiveness of the reasons that anyone in the Burning Hotel example would have, we might be tempted to concede that these reasons could be provided only by objective non-naturalism. However, this decisiveness can also be explained by the natural facts about, and natural similarities between, human beings. Fire is excruciatingly painful to all normally constituted human beings. Nobody likes to endure excruciating pain. Dying will prevent everyone from doing others things that they like to do. Given these natural facts about human beings, the reasons that people would have in Burning Hotel example can be completely explained in a naturalistic manner.
2.6 Conclusion

I have given a small selection of the many, many arguments against naturalistic theories of moral and non-moral value that Parfit raises in *On What Matters*. I have attempted to capture the spirit of Parfit’s enterprise as a whole. I have attempted to examine the implications of his arguments for Railton’s naturalistic theory of moral and non-moral value, and show that Railton’s theory has the resources to weather the storm of Parfit’s objections and still remain plausible. Parfit’s arguments against subjectivism claim that it is unacceptable due to the unintuitive results of various improbable thought experiments. This outcome does not rule out subjective theories, it just shows that their results are a function of their input. His argument from the source of reasons, I think, ultimately shows a virtue of Railton’s theory: that it takes into account both subjective and objective factors of any situation. His arguments against hard naturalism and naturalistic normativity, really show only that naturalism does not provide non-natural reasons, which naturalism never purported to do in the first place. For these reasons, I find Railton’s theory of moral and non-moral value realism to remain a plausible account of value realism.
CHAPTER 3: CONNIE ROSATI

In her article, “Persons, Perspectives, and Full information accounts of the Good,” Rosati argues that theories such as Railton’s, which she calls “Ideal Advisor” views, are implausible. Ideal Advisor views, for Rosati, are views in which one’s good is constituted by what an idealized version of oneself would recommend to the actual non-ideal version of oneself. Rosati argues that Ideal Advisor views of a person’s good fail in at least three ways, (1) they fail to provide the requisite motivational connection between a person and an account of that person’s good, and (2) the addition of full information and rationality are not sufficient traits for an ideal advisor, the advisor must have the correct personality as well, and (3) fully informing someone is a conceptual impossibility. I will begin by examining her account of Ideal Advisor views and normativity. I will then examine her account of Ideal Advisor views and being fully informed. In these first two sections I will be recounting Rosati’s explanation of why an Ideal Advisor view may seem compelling. I will then examine her critique of the view, how and why she thinks that its seeming success is merely apparent. I will conclude that her objections are not insuperable for Railton’s view.


103 Rosati, “Full Information,” pp. 297-299. Rosati lists several metaethical theories of the good for a person which she considers to be Ideal Advisor views, including Sidgwick’s, Rawls’, Brandt’s, and Railton’s. She focuses mainly on Railton’s, which she takes to be the most promising of these.
3.1 Ideal Advisor and Normativity

Rosati begins by noting that Ideal Advisor views do seem to carry normative force, since our Ideal Advisors are better-informed versions of ourselves, and we are people whose opinions we respect.\(^{104}\) Rosati holds normativity to be a conjunction of motivational force for a person and the justification or appropriateness of the object of recommendation. This conception of normativity leads her to provide two conditions that a view of the good for a person must satisfy if it is to provide normativity. The first she calls the internalist requirement.\(^{105}\) The internalist requirement is itself divided into two tiers. The first-tier states that in order for something to be good for a person, A, it must be something that A could be at least capable of caring about. Because people may not care about their good in whatever current state they may find themselves, this condition requires only that A be capable of caring about A’s good under some ideal set of conditions.

To this basic internalist requirement, Rosati adds a second internalist tier.\(^{106}\) There are many possible counterfactual conditions in which a person, A, could find oneself. We do not want to consider A’s judgments in just any conditions to constitute A’s good. Therefore, the conditions that we consider ideal must be further constrained. To maintain the motivational link between A and A’s good, the ideal counterfactual conditions for determining A’s good must be conditions that A finds motivationally compelling as A’s actual self. Again, however, a person may actually be in conditions that are not conducive to caring about counterfactual conditions or theoretical notions of the good. Thus, actual A must find the opinion of A’s ideal self in the counterfactual conditions motivationally

\(^{104}\) Rosati, “Full Information,” pp. 299-300.

\(^{105}\) Rosati, “Full Information,” p. 300.

\(^{106}\) Rosati, “Full Information,” pp. 300-1.
compelling, were actual A to consider them from what Rosati calls ordinary optimal conditions.

The second link must thus be understood to require only that such information motivate a person under "ordinary optimal conditions"—whatever normally attainable conditions we ordinarily regard as optimal for reflecting on judgments about our good. Such conditions include that a person be paying attention, that she be free from emotional distress or neurotic worries, and that she not be overlooking readily available information.\textsuperscript{107}

These are just the conditions in which we normally consider ourselves capable of making serious decisions. This requirement follows from the spirit of internalism, which is that a person’s good must somehow match or fit that person. We intuitively judge that A’s good could not be something that A finds entirely alien.\textsuperscript{108} One way in which A’s good could be alien would be if A could care about it only under some bizarre counterfactual conditions that actual A does not care about or would even find repellent, such as being brainwashed or hypnotized. Unless A finds the opinion of A’s counterfactual self motivationally compelling, the recommendations of this counterfactual self will fail to have normative force.

The second normativity requirement which an Ideal Advisor view must meet, she calls the justificatory requirement.

In addition to their recommending force, judgments about our good have a critical character. We know that our actual desires and aversions can be ill informed, based on mistakes in reasoning, or due, for instance, to

\textsuperscript{107} Rosati, "Full Information," p. 301.
\textsuperscript{108} Rosati, "Full Information," p. 301.
traumatic or atypical experiences. Something that we desire can be good for us, we think, only if that desire is in some sense justified or defensible. The mere fact that someone desires something does not entail that it is part of that person’s good. We all recognize that we are sometimes mistaken and that we occasionally desire things that we recognize later not to have been in our best interest. Therefore, the justificatory requirement says that an Ideal Advisor view must be adequately critical of desires, so as to distinguish a person's actual good from his merely apparent good. When an actual person, A, is asking whether a certain desire is actually or merely apparently good, A is wondering whether A’s consideration of the relevant facts is adequate, and whether A is making any mistakes in rationality. By satisfying the internalist and justificatory requirements, an account of a person’s good succeeds at providing normativity.

3.2 Ideal Advisor and Full Information

An ideal advisor is a version of a person who has been fully informed. But just what does it mean to be fully informed? Such a person must not have simply received all possible information, but must also understand all of the information. Rosati calls the problem of going from mere transmission of information to adequate understanding of information the problem of appreciation. In order for someone to understand information adequately that person must be capable of considering the information in a fully vivid way. As Railton has pointed out, in some cases of information about experiences, this may

---

110 Rosati, "Full Information," p. 304.
mean that a person must actually have undergone experiences in order to adequately and vividly appreciate it.111

When, in the present idealization, it is required that information be fully vivid, it is in effect required that the individual have undergone whatever experience or education would be necessary for this.112

Thus, part of the process of fully informing A will include that A+ has firsthand experience of all of A’s possible options.113 So for the purposes of the rest of Rosati’s discussion, we will suppose that A+ has experienced all of A’s possible lives. These experiences allow A+ to vividly grasp the differences between A’s options and to form informed desires about them.

However, what it means for a person to “vividly” understand information may be a little unclear. Rosati suggests that we interpret the understanding of information so that if further presentations of the same information fail to alter a person’s desires, then that person understands the information vividly.

Let us say that a person has received information in a maximally vivid way when no more detailed representations of it and no representations of it in yet different media or modes of representation would further alter her reactions. That is to say, information has been rendered maximally vivid

111 Rosati, “Full Information,” p. 305.
113 This informing process in which A+ experiences all of A’s possible futures has brought with it the potential for a host of other problems, which are not a part of Rosati’s objections, that I will be unable to address in this work. These include the fact that the order of the experiences will necessarily affect A+’s interpretation of subsequent experiences. It has been suggested that this could be avoided by stipulating that A+ forget all previous idealization experiences, during the experience of particular lives, and having all of the memories restored at the end of the process. See Sobel, David. “Full Information Accounts of Well-Being” Ethics: 1994, 104, 784-810.
when its motivational impact as a function of types of representations has been exhausted.\textsuperscript{114}

Thus, analogous to saturation in chemistry, in which no more of a substance can be dissolved into a liquid, we can say that a person understands information fully when that person has been saturated by that information. That person will be unmoved by further exposure to the information. This vivid and complete understanding of all information truly seems to place ideal advisors in a excellent position to determine the good for a person.

Rosati notes a further merit to this interpretation.\textsuperscript{115} By considering information to be maximally vivid once it has reached its saturation point, we avoid begging the question in favor of a particular view of a person’s good. We would be begging the question if we instead considered information to be maximally vivid only when we got a specific reaction that we prefered to see a person have to information. Instead, the reaction depends solely on the way a person reacts to the full information. Since people differ, different people may react differently to the same information.\textsuperscript{116}

3.3 Rosati’s Objection to Ideal Advisors and Normativity

Although the idea of an ideal advisor as a fully informed version of a person does initially seem to be attractive, Rosati doubts that the opinions of such a person actually provide normativity.\textsuperscript{117} Rosati argues that the problem with an Ideal Advisor view stems from the nature of what it is to be a person with a particular point of view. Sometimes people have psychological features which prevent them from understanding certain forms

\textsuperscript{114} Rosati, "Full Information," p. 306.
\textsuperscript{115} Rosati, "Full Information," p. 306.
\textsuperscript{116} Rosati, "Full Information," p. 307.
\textsuperscript{117} Rosati, "Full Information," p. 307.
of information.\textsuperscript{118} As an example she provides a person who is fully honest, but devoid of empathy, say Betsy. Betsy will tell things as she sees them to be in all aspects of life with no regard to the feelings of others. She will be unable to understand any complaints about how her honest actions negatively affect others. Rosati argues that this is not a problem unique to a lack of empathy; similar limitations will apply for any other person. The current structure of each person's personality limits and determines what information one is capable of absorbing.

Thus simply providing people with unlimited brain power and full information will not allow them to appreciate fully that information if their underlying personalities remain intact.\textsuperscript{119} However, a person's personality may change simply through exposure to new information, especially the saturation of that information. In cases where new information does not alter a person's personality, that person will remain unable to become fully informed about some things. In some cases where a person's personality does change with the addition of new information, the changes to the person may be quite extreme. This is likely to be the case for Railton's Ideal Advisor, A+, who has to consider all the possible futures available to A in order to determine which future is the best. Railton describes A+,

\begin{quote}
Give to an actual individual A \textit{unqualified cognitive and imaginative powers}, and full factual and nomological information about his physical and psychological constitution, capacities, circumstances, history, and so on.\textsuperscript{120}
\end{quote}

\textsuperscript{118} Rosati, "Full Information," p. 308.
\textsuperscript{119} Rosati, "Full Information," p. 308.
Due to the drastic extent an actual individual must change, the ideal advisor will be radically different from the actual individual being advised. Rosati doubts that we should still consider ideal versions of ourselves to be us.\textsuperscript{121} For this reason she thinks that the recommendations of such an ideal advisor may well be alienating for the actual individual, and therefore the normative force of the view is broken.

Rosati considers a possible rebuttal to this objection.\textsuperscript{122} It is true that one’s ideal advisor is quite different from one’s actual self. However, the difference between you and the ideal you is not necessarily problematic. What matters is not that the ideal advisor is still you, what matters instead is that the ideal advisor is still someone whose advice possesses normativity. Since the ideal advisor began as you, and has more information than you, and it is your good that is being determined, perhaps the normativity of the view remains intact.

However, argues Rosati, this rebuttal fails. The qualities that would be required by an advisor in order for that advisor to be ideal are not exhausted by full information and perfect rationality.\textsuperscript{123} Thus, a fully informed and rational advisor does not as such pass the justificatory requirement (that an advisor must be sufficiently critical to distinguish one’s actual good from one’s apparent good). The characteristics of the advisor’s personality are also important. We must begin with a substantive evaluative conception of the personality of an ideal advisor. A substantive evaluative conception of an ideal advisor would be a conception that stipulates that (in addition to merely descriptive properties like being fully informed and perfectly rational) such an advisor would also possess evaluative properties such as being perfectly benevolent, just, kind, etc.

\textsuperscript{121} Rosati, "Full Information," p. 311.
\textsuperscript{122} Rosati, "Full Information," p. 311.
\textsuperscript{123} Rosati, "Full Information," p. 312.
On Railton’s view the personality of an ideal advisor is a function of the initial personality of the actual person and the changes that personality undergoes during the idealization process. Since the addition of information and rationality do nothing to guarantee that the personality of the ideal advisor will be one that renders such a person’s advice authoritative, the ideal advisor view is unsuccessful. Such a person is not actually an *ideal* advisor. The advice of such a person is not authoritative. Instead, a truly ideal advisor must possess the specific type of personality that would make the advice authoritative. Only with the correct personality could an ideal advisor view meet the justificatory requirement. If we consider as ideal, advisors with whatever personalities people happen to have after becoming informed, then people are left hostage to their current personalities and how those personalities happen to evolve through the informing process. Unfortunately, importing a substantive evaluative conception of an advisor’s personality goes against the reductive naturalism that inspired the Ideal Advisor project in the first place (the attempt to provide a theory of the good for a person in purely descriptive terms). However, argues Rosati, importing a substantive evaluative conception of an advisor’s personality cannot be avoided. Either Ideal Advisor views use a substantive evaluative conception of the personality of the advisor, or the views will fail to possess normativity.

Rosati considers two possible responses to her objection that full information and rationality do not alone make an advisor ideal.\textsuperscript{124} The first is to deny her assertion and claim instead that full information and rationality do make an advisor ideal. The second is to claim that if full information and rationality alone do not make an advisor ideal, then any

\begin{footnote}
\textsuperscript{124} Rosati, "Full Information," p. 312.
\end{footnote}
further qualities of personality that may be needed will indeed be brought about through the addition of full information and rationality.\(^{125}\)

Should proponents of the Ideal Advisor view take the first path and claim that becoming fully informed and perfectly rational is all that is needed to make a person an ideal advisor, then Rosati contends that this approach is to begin with a substantive evaluative conception of the ideal personality.\(^{126}\) She calls the view that information and rationality are all that an advisor needs the critical purist view. By maintaining that whatever personality happens to result from the process of adding full information and rationality is sufficient for an ideal advisor, one is already committing to a substantive evaluative conception of personality, namely that whatever personality results is the ideal one. If we are going to begin with an ideal, then that ideal must be justified somehow. To assume that the critical purist ideal yield the best possible advisor is to beg the question against other possible ideals for advisors, such as ones that say they must be benevolent, just, kind, etc.

The second path that proponents of the Ideal Advisor view might take is to claim that the process of becoming fully informed and perfectly rational will alter a person’s personality as required for an advisor’s reactions to be authoritative.\(^{127}\) However, our ordinary experience does not lead us to believe that the addition of information leads people’s personalities in any one direction. Suppose we decide that the ideal personality for our advisor to have is one of benevolence. In our ordinary lives we find that for some people, the more that they learn, the less benevolent they become. Likewise should we decide that the ideal personality is not benevolent but cold and calculating, we often find

\(^{125}\) Rosati, "Full Information," p. 313.
\(^{126}\) Rosati, "Full Information," p. 312.
\(^{127}\) Rosati, "Full Information," p. 313.
that for some people the more that they learn, the less cold and calculating, and the more benevolent that they become. Although Rosati acknowledges that the fully informing process is much more drastic than our everyday experiences, if we are to make reasonable guesses as to what will result from it, we must extrapolate from the actual informing processes we find in the world. Since our experiences with personality changes in the face of increased information do not invariably move in single direction, it seems doubtful that an idealized fully informing process will move toward the personality that yields the best advisor. Thus, the reactions of a fully informed person cannot be taken to be authoritative.

3.4 Rosati’s Conceptual Impossibility Objection to Ideal Advisor and Full Information

The Ideal Advisor view faces an even bigger problem. Due the nature of personhood and the fact that a person must occupy a perspective, it follows that no person could ever successfully become fully informed.\textsuperscript{128} The attraction of the Ideal Advisor view is that, as a person with full information and rationality, A+ will have a perspective that allows A+ to choose between the possible futures of A. However, this idealization is a conceptual impossibility. The personality of A+ will limit the number of potential lives about which A+ could possibly be informed. Recall that in order to become vividly informed about A’s possible futures, A+ must actually experience them during the idealization process. Having undergone the idealization process, A+ will have a certain particular personality. The versions of A that A+ will be when living the possible lives available to A may have different personality types from each other and from that of A+. In order for A+ accurately to evaluate the different possible lives at the end of this process, A+ will need to be able to

\textsuperscript{128} Rosati, "Full Information," p. 314.
accurately recall those lives as they were to each of the non-idealized A’s that A+ was while experiencing those lives. The different conflicting personality types of A+ and the non-ideal A’s of the possible lives will prevent the A+ from being able accurately to recall and evaluate some of the possible lives.

To explain this difficulty, Rosati returns to her example of Betsy, the completely honest person who is devoid of empathy. If Betsy is truly devoid of empathy, then she is by definition unable to imagine how it feels to be someone else. Since the idealization process does not rule out any personality types, it is possible that the fully informed person may end up with the trait of being obtuse. If that is the case, then she will be unable accurately to envision and evaluate all of the possible lives in which her personality is not that of an obtuse person.

Again Rosati anticipates two possible responses to this objection. The first is that it does not matter what traits the fully informed person ends up with. All that matters is that A+’s memory works and A+ can remember all of A’s possible lives and whether and how much A+ liked them while experiencing them. Being able to experience them now in the same way is irrelevant. The second is to reply that the idealization process will in fact make A+ empathetic, because that is what is required to fully understand all of A’s possible lives.

Rosati critiques the first potential response by arguing that the final personality traits of A+ will affect A+’s memory, just as it would affect A’s actual experience of a life.

---

129 Rosati, "Full Information," pp. 317-318. Rosati has a non-standard use for the word ‘obtuse’ and uses it exclusively to mean devoid of empathy and unable to comprehend the feelings of others.
130 Rosati, "Full Information," p. 319.
132 Rosati, "Full Information," p. 320.
If one cannot now experience life in a certain way, then it is also impossible for one accurately to remember experiencing it in that way.\footnote{Rosati, "Full Information," p. 320.}

If she cannot now see another’s distress as calling for a sympathetic response, it is unclear how she could accurately remember what it was like to so view it. In order to recall what it is like to see another’s distress as calling for a sympathetic response, it seems that she must now be capable of seeing another’s distress in that way.

This is not a problem exclusive to the lack of empathy. It is equally applicable to other conflicting traits such as trustfulness and distrustfulness, or having a sense of humor and having no sense of humor. Rosati argues that if A’s fully informed self ends up distrustful, A+ will be unable accurately to remember life as a trustful person, or if A+ ends up without a sense of humor A+ will be unable accurately to recall what it is like to find things funny.

The second possible response from proponents of an Ideal Advisor view is to amend the idealization process to include adding full information, rationality, and empathy. If empathy is the ability to understand what other people are thinking and feeling, then perhaps if the ideal advisor were fully empathetic, A+ would be able to know and understand what it is like to be any of A’s possible selves. To this Rosati argues that our ordinary notion of empathy is still insufficient to allow A+ to appreciate fully all of A’s possible lives. Empathy allows a person non-judgmentally to understand other people’s thoughts and feelings by drawing on one’s own experiences.\footnote{Rosati, "Full Information," p. 322.} Empathy does not entail that A+ actually knows what it would be like to be that other person.

Rosati objects that if we attempt to stipulate that the fully informed person will have an ideal empathy that does allow A+ to know what it is like to be other people, then this
notion of ideal empathy will still be mysterious and problematic. Worse still, if A+ does have the empathy needed to understand all perspectives, then A+ will no longer have a particular perspective from which to form preferences about different lives. A+’s personality traits were what allowed A+ to form such preferences. If A+’s personality is so dilute as to allow the understanding of all perspectives, A+ will not be able to prefer some lives to others.

3.5 Does Railton meet the Internalist Requirement?

Let us consider the first possible problem Rosati raises for an account of the good for a person. The good it describes must be capable of motivating the person whose good it is, at least in some circumstances. This connection is what Rosati attempts to address with her internalist requirement. How does Railton’s theory fare on this matter?

By Railton’s own lights, an account of the good for a person must be suitable for that person, rather than alienating to that person.

It does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone’s good to imagine that it might fail in any such way to engage him.\(^{135}\)

It is reasonable to expect Railton’s view to meet at least some version of the internalist requirement, because he imposes it on himself. At first glance it seems that Railton’s theory may have been designed specifically to meet this problem. On Railton’s theory, not only is the good for a person in “some way” suited to that person, it is constituted by the

\(^{135}\) Railton, “Facts and Values,” p. 47.
desires of an idealized version of that person. We know that A’s good is capable of attracting A, because *ex hypothesi*, a A’s good *is* what attracts an idealized version A.

Recall that Rosati’s internalist requirement was in the form of two-tiers, a minimal tier and a more demanding tier. The minimal tier is just that one’s good must be capable of motivating one under “ideal conditions.” The previous paragraph shows that Railton meets the requirement of this tier. The second-tier is that the ideal counterfactual conditions for determining A’s good must be conditions that A finds motivationally compelling as A’s actual self. The need for a second-tier is intuitively plausible. There are many possible counterfactual conditions under which A could contemplate A’s good. It may seem that we do not want the counterfactual conditions which determine a person’s good to be alienating either. Thus, Rosati requires that the counterfactual conditions under which an ideal advisor determines a person’s good be motivationally compelling to the actual non-idealized person, when considered by the non-idealized person under “ordinary-optimal conditions” (sober, calm, composed, and considering available information).

Although it may be controversial as to how often Railton’s counterfactual conditions are found motivationally compelling, I do not think that this controversy implies that Railton’s theory of the good for a person fails to apply to everyone. Since this claim is incompatible with Rosati’s second-tier, I will now provide an argument against the second-tier. The reason why actual people are often unable accurately to predict their own best interests is due to the inherent limitations of actual human beings. As actual people, we are ignorant, and often fail accurately to weigh or consider all of the relevant information. These limitations remain in place even during ordinary optimal conditions. Thus, two-tier internalism seems to be a highly questionable hypothesis. It seems that some actual people may not care about their good even in their most calm, lucid, and rational
moments. People indifferent to their good, or people with misguided conceptions of their good, might not care about the opinion of any hypothetical advisor, even if it is an ideal version of themselves. However, I would not be willing to say that an actual person who does not care currently about his good even in optimal conditions has no good to speak of. A fully informed and rational version of such a person, A+, might well want A to begin to care about A’s good. Thus, although simple one-tier internalism seems plausible, the second-tier of Rosati’s internalism is an unreasonable demand.

To illustrate the failure of two-tier internalism let us consider the case of Jack, a primitive nomadic tribesman. Jack has a cut on his torso that has become infected. He consulted his tribe’s witch doctor who recommended that he catch a rare desert snake to make into a medicinal stew that will cure him. Jack is now in the desert waiting for a snake to pass by and is calm and collected and has complete confidence in his full recovery. He has no worries and is in what for Jack are ordinary optimal conditions for considering questions about his good. Suppose Jack were asked if he would consider the counterfactual conditions of a fully informed and perfectly rational version of himself to be motivationally compelling, as an account of his good. “No”, says Jack, “the witch doctor is the authority on my good. Even under such ideal conditions, I would not know my own good.” Further suppose that Jack’s tribal treatment will not cure his infection, and will lead him to die an agonizing but preventable death. Railtonian-idealized Jack, Jack+, would recognize that Jack was suffering from a bacterial infection that could be completely cured with antibiotics available in the local village. Jack+ would want Jack to want to procure the antibiotics from the village and to take the full course of them, which would lead to his complete recovery. The mere fact that actual Jack is incapable of recognizing or being
motivated by the advice of Jack+ should in no way disqualify those conditions from being a part of the account of the good for this person.

Just as there is often a major discrepancy between actual people’s motivation and their good, a discrepancy often occurs between people’s motivation and theories of their good. Given these considerations and many other easily imaginable and plausible counterexamples in which an actual person, even in ordinary optimal conditions, is in no position to recognize conditions for an account of his good, I find the second-tier of Rosati’s two-tier internalism to be an unacceptable requirement. The mere addition of true information and flawless rationality is not alienating in the way brainwashing by an evil scientist or other bizarre counterfactual conditions might be. 136 The only relevant internalist requirement is that a person’s good be suited for them, which is the case if it is capable of being found attractive by an idealized version of themselves. Railton’s theory delivers this motivational connection.

3.6 Railton and the Justificatory Requirement

Rosati’s justificatory requirement is a much more significant hurdle for Railton’s theory to overcome. The justificatory requirement says that an Ideal Advisor view must be adequately critical of desires, so as to distinguish one’s actual good from one’s merely apparent good. Rather than asking whether actual agents will find the desires of their idealized selves motivationally compelling, we must ask whether they should find them motivationally compelling. Rosati suggests two ways in which Railton’s theory may fail the justificatory requirement: (1) full information and rationality are not sufficient traits for an

136 By alienating, I mean that most people would fail to find it motivationally compelling. I take it to be the case that most people would see the attractiveness of the desires of fully informed and rational versions of themselves.
ideal advisor, the advisor must have the correct personality as well, and (2) fully informing someone is a conceptual impossibility.

3.6.1 Ideal Advisor’s Personality

The attractiveness of the Ideal Advisor view lies in its creation of an objective perspective (a fully informed and rational perspective) from which to determine the good for a person. Rosati objects that the traits of being fully informed and perfectly instrumentally rational do not exhaust the traits that are needed in an advisor who is ideal. In other words, the advice of versions of ourselves who are fully informed and rational will not necessarily be advice that reflects the good for us. In some cases the personality of the individual, even when idealized in this way, will not yield a person whose desires should be held as authoritative. In rejecting the justificatory power of Railton’s A+’s fully informed desires for A, Rosati argues that we require a substantive evaluative conception of an ideal person in order to have reason to believe that an ideal person A+ could truly determine a person’s good. In other words we must first know what qualities an ideal advisor would have before we can decide whether some counterfactual conditions would yield an advisor that is ideal. But why should we think this to be the case?

Rosati attempts to defend this assertion with the story of Sandy.¹³⁷ Sandy has one type of personality, and is wondering if she should change her personality to become a better person with a different personality type. Rosati argues that since Sandy is contemplating changing the type of person she is, she may not and maybe should not find the advice of an ideal version of herself compelling, since it will be an ideal version of herself who is currently not the type of person she wants to be. Her idealized self will still

be just a person with her current flawed personality type and thus be making the choice from a perspective that Sandy does not endorse.

This situation is not a problem that a Raltonian idealized version of Sandy would be unable to handle. Consider how actual Sandy would attempt to determine the type of personality that she should cultivate. She will do this by surveying the different people she knows (possibly seeking out additional role models in her society, or literature) and then attempting to decide which of these personalities seem attractive to her. She may look at how these people behave differently from how she behaves in similar situations. She may look at how happily their lives are going, and how their personalities affect other people’s lives, or how their personalities affect their outlooks on the world in general. From the information that she is able to gather on different personality types, she will attempt to decide which of these lives would be a good fit for her based on how attractive she views the projected outcome of developing the different personality types (and the good she would do for others may be a component of this attractiveness).

Sandy+ will share Sandy’s concern of what personality type she ought to cultivate, because Sandy+ is contemplating Sandy’s good as a person about to enter Sandy’s situation. Therefore, Sandy+ will be facing the choice of whether to maintain Sandy’s current personality or attempt to cultivate a new one. The difference is that rather than deciding from a limited perspective, Sandy+ is vividly aware of all of the possible personalities available to Sandy. She will also know all of the actual rather than expected consequences of cultivating a new personality. The position of Sandy+ is the best position, in principle, from which to determine what personality is right for Sandy.

Rosati worries that rather than simply have an advisor with full information and rationality, that we ought to have a substantive evaluative personality for our ideal advisor.
This idealization might mean that in addition to full information and rationality, benevolence would also be added. However, suppose that we similarly idealize Sandy, but with the stipulation that she have the personality trait of benevolence, making Sandy++. Then from her ideal standpoint Sandy++ might desire that Sandy desire a different personality from what would be desired by Sandy+. Suppose that Sandy++ would desire that Sandy cultivate a much more altruistic lifestyle than the one that would be desired by Sandy+. The problem is that rather than Sandy with additional benevolence living the altruistic life, it will be Sandy who does not possess this additional benevolence. Sandy could be at best equally pleased with her new personality than with the personality selected by Sandy+ (if it were the personality most desirable for actual Sandy, then Sandy+ would have selected it). However, we now face the possibility that Sandy++, given her potentially foreign\textsuperscript{138} trait of benevolence, selected an option that will be less good for Sandy, because Sandy++ made her decision from a position of additional benevolence which may have valued the consequences of Sandy’s personality for others higher than the consequences of the personality for Sandy. This discrepancy between the good desired by Sandy++ and Sandy+ is not to suggest that the life selected by Sandy+ will be excessively selfish, only that it will be selected from a position of desiring the things that Sandy would desire if she were vividly aware of all the facts. Sandy+ may also become benevolent upon knowing the facts, but if this is the case her benevolence is the result simply of her understanding the relevant information, not because we stipulated that Sandy+ must also become benevolent.

\textsuperscript{138} The benevolence added to Sandy++ is foreign if it is a trait that Sandy would not have developed simply through receipt and understanding of all of the relevant information. If Sandy would develop this benevolence simply by through receipt and understanding of all of the relevant information, then Sandy++ = Sandy+ and there is no disagreement about Sandy’s good.
If we reject the second-tier of Rosati's internalism, we need not worry that Sandy+ is unable to determine Sandy’s good simply because actual Sandy may doubt Sandy+’s advice. Actual agents may fail to be motivated by theories of their good just as they may fail to be motivated by their actual good. However, we still want Sandy’s good to be suited for Sandy, rather than alienating. It is interesting to note that Rosati is the one attempting to foist questionable personality traits upon the ideal advisor. Given full information and rationality Sandy+ either would or would not have the personality trait of ideal benevolence. If she did then Sandy+ and Sandy++ would be the same person and we would have no dispute. However, if Sandy+ did not possess ideal benevolence, then why would Sandy find the advice of a person with personality traits that she would not have if she were aware of all of the facts compelling?

Rosati considers two possible responses to her personality objection. The first is to deny her assertion and claim instead that full information and rationality do make an advisor ideal. The second is to claim that if full information and rationality alone do not make an advisor ideal, then any further qualities of personality that may be needed will indeed be brought about through the addition of full information and rationality. These are similar and my response can be seen as a combination of the two. Yes, full information and rationality make an advisor ideal, however, becoming fully informed and rational may change the qualities of one’s personality. Remember the idealization process holds intact all of person’s non-belief properties, other than those that change naturally through the process of gaining all true beliefs about everything relevant to one’s desires and being capable of vividly contemplating these facts.

---

139 Rosati, "Full Information," p. 313.
Rosati has rebuttals to each of these potential responses. First she notes that if you assume that full information and rationality will make an ideal advisor, come what may, then you are actually invoking an evaluative conception of personality, one that is fully informed and rational. This claim seems plainly false. If you add evaluative personality traits to an ideal advisor, such as benevolence, or kindness, or charity, then this is to begin with an evaluative conception of personality. If I refuse to import value judgments into a descriptive account of the good, then I am not beginning with an evaluative conception of personality. This ideal advisor can be described with purely descriptive terms. If Rosati insists that the absence of an evaluative conception of personality is itself an evaluative conception of personality, then it is impossible by definition to have a descriptive account of the good. This requirement is question begging against the naturalist who argues that the good for a person can be accounted for purely descriptively.

Rosati also has a rebuttal to the response that any necessary differences in personality will come about through the process of fully informing an individual. She asks us to consider some additional personality traits, for example, benevolence, patience, and kindness. However, we have no reason to believe that the idealization process would invariably bring these about. In our actual lives we find that more information and experience leads some people to become more cynical and cruel instead. This is true regardless of what traits we choose. Ordinary experience shows us that more information and experience do not always tend to promote any single personality trait or set of personality traits. However, this response forgets or misunderstands that we are

140 Rosati, “Full Information,” p. 313.
attempting to provide a theory of the good for a person. Each person need not gain the
same traits as everyone else.

Rosati is correct in arguing that in order for advisors to be ideal they need to have
the correct set of properties. However, she goes wrong by insisting that we can begin with
a substantively ideal advisor. We cannot agree beforehand on an evaluative conception of
personality and insist that all advisors end with these traits. People are different, and we
cannot know a priori that any single personality would be fitting for all advisors. Instead
ideal advisors gain the correct traits through a methodological idealization process. By
adding a complete understanding of the relevant facts and perfect rationality, we are able
to create advisors who are ideal at determining the good for their non-idealized selves.
Railton’s theory does not yield results on what is good for most people, or what is good for
people in general. It is a theoretical account of the good of any given person, and it does
this by proposing that each person’s good can be determined by idealizing that person. It
is not an idealized version of someone else who determines what is good for you.

One of the purposes of Railton’s reductive naturalistic project of the good for a
person is to explain how a person’s good needs to be suited to that person. This is the
spirit behind Rosati’s own internalist requirements. The good for a person is a non-moral
value. If a person (S) is inherently selfish, it is unlikely that an idealized version of him S++
with the trait of benevolence stipulatively added would necessarily desire what is in the
best interest of S’s selfish actual self. Of course if S’s selfishness is in any way grounded
in false beliefs, such as mistaken predictive outcomes of selfish behavior (such as that
selfishness benefits oneself more than selfishness hurts oneself), then simply fully
informing him will be enough to show S+ the error of S’s ways. If S+ has a complete
understanding of the consequences of S’s selfishness, and would still desire that S continue in his selfish ways, then selfishness is actually in his best interest.

If adding full information and rationality to a person, A, making A+ would cause A+ to become benevolent, then A+ would be benevolent. If adding full information and rationality would cause A+ to become selfish, then A+ would become selfish. Either of these A+’s might desire that A desire to become benevolent or selfish. A truly selfish A+ might desire that A desire to become truly benevolent. If A+ saw that the A’s life would be most enjoyable by becoming truly benevolent (not just pretending to be benevolent when it suits A) then A+ might desire that A desire to become truly benevolent. The point is that it is facts about A that determine the good for A. An A+ with only full information and rationality will be more like A, and more suited to determine the good for A, then an A++ with foreign personality traits foisted on.

3.6.2 Railton and the Fully Informed Person

Even if all of her previous objections fail, and a fully informed person would yield an ideal advisor, Rosati has still another, stronger objection at her disposal. The fully informed person is a chimera: being a person with a perspective precludes the possibility of being fully informed. If Rosati is right and it is conceptually impossible to be fully informed, then Railton’s theory of the good for a person is unsuccessful.

Rosati understands that we cannot fully inform an actual person for nomological reasons.

But we cannot, as a practical matter, fully inform a person. Moreover, no actual person could be fully informed, it appears, without violating the laws of psychology and physiology.
Her objection is much more than a nomological problem. Instead she argues that the fact that a person must occupy a perspective conceptually entails that no person could ever successfully become fully informed. Rosati objects that at the end of the idealization process, the “fully informed” individual, A+, will have some specific set of personality traits. These traits are what allows A+ to prefer some possible futures for A. Since the possible futures of A include all futures nomologically available to A, A’s possible futures will include lives in which A develops many different personality traits. A+ has only one fixed set of personality traits, and has to be able to contemplate and vividly compare all of the possible lives of A. Since there are likely to be possible lives in which A develops personality traits which are contrary to A+’s personality traits, A+ will be unable to accurately imagine or remember what it is like to be versions of A that are contrary to A+’s personality. Thus, A+ is unable to be fully informed about these possible lives. So, A+ is unable to determine the good for A, and Railton’s theory fails.

Rosati considers two potential responses to this objection available to us.\textsuperscript{141} The first is that it does not matter what traits the fully informed person ends up with. All that matters is that A+’s memory works and A+ can remember all of A’s possible lives and whether and how much A+ liked them while experiencing them. Being able to currently experience them the same way is irrelevant. The second is to reply that the idealization process will in fact make A+ empathetic, because that is what is required to fully understand all of A’s possible lives.

As before I will attempt to defend both of these responses. Let us consider the first: A+ has an ideal memory. Even if A+ cannot currently be someone with conflicting personality traits, A+ can remember what each of these separate lives was like, and so

\textsuperscript{141} Rosati, "Full Information," pp. 319-24.
compare them to each other now as A+’s current self. This possibility seems to be intuitively plausible. However, Rosati has a rebuttal to this response. Memory is a type of experience. Since A+ cannot currently experience both types of lives with A+’s current personality, so too A+’s current personality traits will affect the accuracy of A+’s memories as well.

Our own experiences indicate that while we can sometimes recapture perspectives we occupied at previous times in our lives, often we are quite unable to do so. We may be unable, for instance, to recapture the perspective we had before undergoing an important kind of training or therapy, or before we fell in or out of love for the first time, or before we experienced a particular trauma or revelation. Our inability is not due to insufficient memory or imagination but to the character of the change that we have undergone.

Thus, A+’s ideal memory is not a solution. A+ will have a personality, and having a personality prevents one from remembering what it is like to have personalities that conflict with one’s current one. This inability is demonstrated by contemplating our own experiences.

The problem with this rebuttal is its mistaken account of how our memories and experiences work. I think we can remember what it was like to have different personality types that conflict with our present personality. If we cannot then it is due to insufficient memory or imagination. Surely we have experienced what it is like to have our personalities change and develop over time as we have become older and wiser. We can all look back and reminisce about what it was like to be a young dumb kid. We can see that our personalities have changed. Hopefully most of us think that we have changed for
the better. Some people may remember what it was like to be innocent and long for those carefree days of the ignorant past, thinking those times actually better. Very few people remember their pasts perfectly accurately. But this inability is due, I suggest, merely to the limited powers of the memories of most human beings. Giving us unlimited cognitive power would solve this problem. Whatever our fully informed personalities and perspectives turn out to be, I see no reason to believe that our idealized selves would not be capable of remembering our possible lives with complete accuracy, and therefore authoritatively choosing between them from their idealized perspectives.

This conclusion, I believe, is supported by our experiences of remembering how differently we viewed the world before certain major personality changes. I think that most people can remember what it was like to have a different personality, one that was in some ways contrary to one’s current personality. We can remember bad decisions that we made and think of what different people we were in the past. This success is all the more likely to be true given that the personality differences between A+ and the different A’s will all be those differences capable of being experienced by the same individual. These personality changes are mediated simply by the addition and vivid understanding of information. Those of us who have undergone major changes in our belief systems and worldviews, to or from theism say, can vividly remember what it was like to have different and conflicting personality traits. Definite successes such as these suggest that our occasional failures to recapture a perspective are not due to a conceptual impossibility. Those of us who cannot remember such differences, I suggest, have either not experienced such changes, or are suffering from a lack of memory or imagination.

The second possible response of a defender of an ideal advisor view is that the idealization process of adding to a person full information and rationality will invariably add
empathy, and thus allow A+ to accurately weigh all of A’s possible lives. This response too has intuitive plausibility when we consider that if empathy would be required to vividly understand the information (what it is like to experience different possible lives) and we are stipulating that A+ *does vividly understand all of the information*, then it follows that A+ must possess whatever empathy is so required. Rosati further concedes that adding empathy to the idealization process is not to import a substantive value judgment as would importing benevolence.¹⁴²

Again Rosati has a rebuttal prepared for this response. She claims that our ordinary notion of empathy is insufficient to do the work we are here asking of it (vividly understanding possible lives in which A’s personality conflicts with A+’s current personality). Ordinarily by empathy...

> We seem to mean instead that she understands another’s thoughts and feelings, possibly drawing on her own similar experiences, and perhaps also that she responds in an uncritical way.¹⁴³

So rather than allowing one to understand fully what it is like to *be* another person, empathy simply provides an understanding of that person’s thoughts and feelings. This characterization seems to me to miss the mark. If empathy allows one to understand someone else by “possibly drawing on her own similar experiences,” then this capacity is all that is needed considering that A+ has *actually experienced* what it is like to be each A. Drawing on past experiences is all that A+ needs.

Rosati next argues that even if one has the requisite notion of empathy, then one will no longer have a perspective from which to choose.¹⁴⁴ All personality traits which one

---

¹⁴³ Rosati, “Full Information,” p. 322.
would use to prefer one scenario to another would now be gone. Therefore, if A+ is sufficiently informed to make an educated preference between A’s possible lives, A+ will have no basis to discriminate between them. This conclusion, I suggest, need not follow considering our notion of empathy (drawing on one’s own experiences to understand others). Suppose that A+ has a certain personality X and is choosing from lives with conflicting personalities G, H, and I, as well as compatible personalities T, U, and V. A+ uses empathy to draw on the actual experiences of A’s lives with personalities, G, H, I, T, U, and V to remember what it was like for these different versions of A while A was living these lives. A+ will not be biased by A+’s current personality X, because while living the life A+ prefers A will not have personality X, A will have one of the personalities G, H, I, T, U, or V. Yet A+ still has a basis to choose between these lives; A+’s vivid understanding of what it was like to be the people living them.

Given the conceptions of empathy and memory that I have suggested A+ will have everything that might be needed to choose between lives. Rosati suggested that A+’s personality would have to be effaced in order to accurately remember A’s possible lives. If it were not, then A+ would be unable remember lives in which A’s personality conflicts with A+’s. In which case A+ would no longer be able to prefer one life to another. Whether or not it is conceptually possible to fully inform a person is a difficult metaphysical question. Due to the abstract conceptual nature of the idealization process, it is very difficult to come to a conclusion on this issue with a strong degree of confidence. This objection may be the strongest one facing Railton’s theory. However, as Rosati notes, if we are to make reasonable guesses as to the results of the idealization process, we must extrapolate from what we know about actual humans. I have argued that we can remember perspectives from our pasts when we had personalities which conflict with our present ones. If some of
us have this ability, then all of our ideal selves undoubtedly would. For these reasons, A+ will have a personality and yet still be able to desire the best possible life for A.

3.7 Conclusion

I have examined three of Rosati’s objections to Railton’s Ideal Advisor view of the good for a person. First, Rosati objected that an ideal advisor must have those traits that would be found motivationally compelling by an actual individual in ordinary optimal decision making circumstances. I argue that this requirement is too strong, and that a view of the good for a person need not meet it. Second, Rosati objected that an ideal advisor view requires us to begin with a substantive evaluative view of personality. This I have argued is not the case. Descriptive changes to a personality need not also be evaluative. Further, I argued that in fact adding an evaluative conception to an ideal advisor could only at best match a non-evaluative advisor, and more likely be inferior. Third, Rosati objected that an ideal advisor cannot both be fully informed and have a perspective from which to prefer some possibilities to others. This too seems to be mistaken. I think that our ordinary experiences support the opposite conclusion.
4: WHAT HAVE I DONE?

In Chapter 1, I explicated a view of moral realism that I wanted to defend. This view was Railton’s reductive naturalistic moral realism. The reductive naturalism of this view is found in the fact that he provides an account of the good for a person (non-moral value) which can be explained entirely in descriptive terms. The good for A is what a version of A that was idealized with full and vivid understanding of all relevant empirical information and perfect instrumental rationality, A+, would want A to want were A+ to enter the non-idealized position of A. Moral Value is then constituted by maximizing non-moral value from a social point of view (one impartial to the interests of particular individuals).

This view of moral and non-moral value is realistic because the values in question are mind independent, and we can interact with them casually. The postulation of these values has explanatory uses. Non-moral value explains why some choices turn out poorly for people while other choices turn out well. It can sometimes explain the evolution of some people’s subjective interests over time. Sometimes as people interact with the world they gain experience and they come to desire things that better match their objective interests. Actions that have bad results are avoided while actions with good results are repeated. Moral value can often explain why some groups of people tend to do well and flourish, while other groups of people do poorly and languish. These outcomes are the results of either approximating or failing to approximate rationality from a social point of view. Groups who better approach this type of rationality tend to cooperate more and thus succeed. Those who fail to approach this rationality are more likely to impede each other's interests, leading to turmoil.
I find this to be an intuitively plausible as well as rationally defensible account of value. Next I went on to examine some objections to the view that were raised by Parfit and Rosati. I ultimately argued that Railton’s theory is capable of overcoming these objections.

### 4.1 Parfit

In Chapter 2 I examined some views of Derek Parfit. Parfit, like Railton, is a moral realist. However, Parfit subscribes to a non-naturalistic stripe of realism that he calls Rationalism. According to this theory people have irreducibly normative reasons to do some things rather than others. Parfit views naturalistic moral theories such as Railton’s, which do not invoke irreducibly normative reasons, as very close to nihilism. According to Parfit in order for things to matter there must be irreducibly normative reasons. This perspective leads him to critique naturalistic theories on a variety of fronts.

Parfit’s first set of objections is aimed against the subjectivist aspects of naturalism. Railton’s theory of non-moral value or the good for a person reduces a person’s objective interests to what an idealized version of that person would want the non-idealized person to want. Therefore, for Parfit, what Railton’s theory suggests that we have reasons to do is unacceptably tied to us as subjects. He illustrates this problem with subjectivism by using a variety of examples meant to show the absurdity of subjectivism. If my idealized self wants me to want to smoke, then I should want to smoke. If my idealized self wants me to be in agony, then I should want to be in agony. I argued that these objections do not succeed. We have no reason to think that when idealized, people like us would want such ridiculous things. Parfit responds that it does not matter if this objection would not apply to any actual people. All that matters is that we can imagine it being the case that someone would want them. In this case, if we are forced to imagine someone who is so different
from us that they could desire to be in agony, then it should no longer be surprising that
agony could be a part of the good for such a strange person.

Parfit’s next objection is that due to the source of the naturalist’s reasons (the good
for the person which is constituted by idealised desires), these desires are incapable of
providing us with reasons. However, in this case, by reasons Parfit means irreducibly
normative reasons. To demand irreducibly normative reasons from the naturalist, who is
not attempting to provide us with irreducibly normative reasons is an unacceptable
demand.

Parfit next objects that hard naturalism, which is the view that value is completely
reducible to natural properties, can yield only trivial theories. A hard naturalistic theory will
be trivial if it either (a) fails to have normativity, or (b) fails to be significant (something over
which we could disagree or something that is informative). Parfit argues that hard
naturalistic theories fail to provide normativity but fails to consider normativity composed of
anything other than irreducibly normative reasons. Naturalism can provide normativity.

Consider an instance when the utterance, “You ought to feed this hungry child,” is true.
This ought can be translated into the completely naturalistic facts of the world (1) If you do
not want the child to suffer, then you must feed her, and (2) you do not want the child to
suffer. These naturalistic facts of the world make the utterance of the normative statement,
“You ought to feed this hungry child,” true. He makes the further claim that naturalistic
theories are non-informative by equivocating between moral and pragmatic instances of
normative terms. When considered without this equivocation naturalistic theories are
informative.

The last of Parfit’s arguments that I considered was his Normativity Objection. This
argument claims that though naturalism may provide us with reasons, these reasons are
of a metaphysical difference in kind from what reasons need to be (irreducibly normative). It may be true that naturalistic and irreducible reasons are of a metaphysical difference in kind. However, the naturalist's reason are able to account for the normativity required by our theories of value. Given this success, we should not rule these reasons out on metaphysical grounds simply from what we currently know about normativity.

Although I ultimately found Parfit's arguments against naturalism unconvincing, I am sympathetic to his plight. Since he believes both (a) that things matter, and (b) that if things matter we must have irreducibly normative reasons, I understand why he believes that we must have irreducibly normative reasons. I agree with him on, and strongly support, (a). As human beings we know that things matter. Any realistic theory of value must account for this fact. (b), however, I reject. It seems to me that a reductive naturalistic theory of value such as Railton's is able to sufficiently explain why things matter. Given the fact that Railton's view also explains why things matter, I prefer Railton's view because I find it to be both a better explanation and a more plausible explanation.

4.2 Rosati

In Chapter 3, I examined some of Connie Rosati's views. Rosati's objections are much more specific to Railton's theory than Parfit's objections were. Rosati's objections are aimed at ideal advisor views (a person's good is constituted by what that person desires under ideal conditions) in general, and at Railton's theory in particular. The objections are specific to his account of non-moral value or the good for a person. Rosati, like Parfit, holds reservations about the normativity of the account. She divided her concerns about normativity into two requirements, the internalist requirement and the justificatory requirement. I argued that Railton's theory is capable of overcoming all of the problems associated with these.
Rosati’s internalist requirement is divided into two-tiers. The first tier is that a person’s good must be potentially motivating for that person, at least under some conditions. Railton’s theory meets this requirement by defining the good for a person A, as what an idealized version of A, A+, wants A to want. The second tier requires that A be motivated by the idealization conditions of A+, if A is in ordinary optimal conditions. I rejected the requirement of the second tier on the grounds that A may plausibly fail to recognize the correct idealization conditions, even were A in ordinary optimal conditions.

Rosati suggests two ways in which Railton’s theory may fail the justificatory requirement: (1) full information and rationality are not sufficient traits for an ideal advisor, the advisor must have the correct personality as well, and (2) fully informing someone is a conceptual impossibility. Rosati argued for (1) by making the case that in order for an advisor to be ideal we must first decide on all of the evaluative personality traits that we want our ideal advisors to have at the end of the process. I argued that this requirement is a mistake. Rather than beginning with a substantively ideal advisor we need only a methodologically ideal advisor. The facts about each particular person will determine the personality traits that are required by an advisor ideal for that person. This idealization will render each advisor maximally effective at determining the good for the specific person being advised. People are simply too different for us to be able to determine a priori the evaluative traits that all advisors will need. Due to the similarities among people, many advisors will likely end up with similar personalities. However, we are in no position to stipulate beforehand what all of these traits will be.

Rosati argued for (2) by making the case that people with some types of personalities are incapable, as a matter of conceptual necessity, of being fully informed about different lives where a person’s personality conflicts with the advisor’s current
personality. She supported this claim by appeal to our actual experiences. I disagreed that our experiences support such a conflict. We are able to remember our previous perspectives after undergoing personality changes. This capacity is all the more likely when we consider that the personality changes are mediated merely through changes in belief and experience. Since some actual people are capable of such memories this is not a conceptual impossibility. Due to the superior cognitive capacities of ideal individuals, I suggest that they will have no problem remembering past perspectives.

I examined each Rosati’s objections in turn and found them each lacking. In so doing I conclude that Railton’s theory remains a plausible account of the good for a person. It will most likely always be impossible to rule definitively that it is successful. We are likely to find many advances in our future enquiry of metaethics. However, I find the current condition of Railton’s theory to remain a highly attractive account of moral realism and the good for a person.
BIBLIOGRAPHY


